

NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
SCHOOL OF SCIENCES
FACULTY OF BIOLOGY

**Managing biodiversity information:
Development of a virtual repository**

Sarah Faulwetter

A thesis submitted for the requirements of the Doctoral Degree



ATHENS 2014

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

**Διαχείριση πληροφορίας βιοποικιλότητας:
Ανάπτυξη εικονικού αποθετηρίου**

Sarah Faulwetter

Διδακτορική Διατριβή



ΑΘΗΝΑ 2014

NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
SCHOOL OF SCIENCES
FACULTY OF BIOLOGY

**Managing biodiversity information:
Development of a virtual repository**

Sarah Faulwetter

A thesis submitted for the requirements of the Doctoral Degree

ADVISORY COMMITTEE

ARTEMIS NICOLAIDOU
(Professor – Supervisor)

ANASTASIOS LEGAKIS
(Associate Professor)

CHRISTOS ARVANITIDIS
(Senior Researcher)

EXAMINATION COMMITTEE

ARTEMIS NICOLAIDOU
(Professor)

ANASTASIOS LEGAKIS
(Associate Professor)

CHRISTOS ARVANITIDIS
(Senior Researcher)

MARTIN DÖRR
(Research Director)

CHARITON CHINTIROGLOU
(Professor)

THEODOROS KEVREKIDIS
(Professor)

ALBERTO BASSET
(Professor)

ATHENS 2014

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

**Διαχείριση πληροφορίας βιοποικιλότητας:
Ανάπτυξη εικονικού αποθετηρίου**

Sarah Faulwetter

Διδακτορική Διατριβή

ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ

ΑΡΤΕΜΙΣ ΝΙΚΟΛΑΪΔΟΥ
(Καθηγήτρια — Επιβλέπουσα)

ΑΝΑΣΤΑΣΙΟΣ ΛΕΓΑΚΙΣ
(Αναπληρωτής Καθηγητής)

ΧΡΗΣΤΟΣ ΑΡΒΑΝΙΤΙΔΗΣ
(Ερευνητής Β')

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

ΑΡΤΕΜΙΣ ΝΙΚΟΛΑΪΔΟΥ
(Καθηγήτρια)

ΑΝΑΣΤΑΣΙΟΣ ΛΕΓΑΚΙΣ
(Αναπληρωτής Καθηγητής)

ΧΡΗΣΤΟΣ ΑΡΒΑΝΙΤΙΔΗΣ
(Ερευνητής Β')

MARTIN DÖRR
(Ερευνητής Α')

ΧΑΡΙΤΩΝ ΧΙΝΤΗΡΟΓΛΟΥ
(Καθηγητής)

ΘΕΟΔΩΡΟΣ ΚΕΒΡΕΚΙΔΗΣ
(Καθηγητής)

ALBERTO BASSET
(Καθηγητής)

ΑΘΗΝΑ 2014

Table of Contents

Table of Contents	i
Declaration	iii
Abstract	v
Περίληψη	vii
Acknowledgements	xi
Preface	xiii
Chapter 1 Introduction	1
Chapter 2 <i>Sphaerosyllis levantina</i> sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in <i>Sphaerosyllis hystrix</i> Claparède, 1863.	15
Chapter 3 Micro-computed tomography: Introducing new dimensions to taxonomy.	37
Chapter 4 <i>Polytraits</i> : A database on biological traits of marine polychaetes.	85
Chapter 5 Comparison of structural and functional stability of polychaete assemblages in coastal lagoons.	129
Chapter 6 Discussion and conclusions	157
Glossary of terms	175
Annex I Penev L, Blagoderov V, Georgiev T, Rycroft S, Scott B, Faulwetter S (2011) ViBRANT deliverable 6.1: XML markup tools and service. 10pp.	187
Annex II Documentation of the data entry interface of the <i>polytraits</i> database	199
Annex III Supplementary material of Chapter 5	211

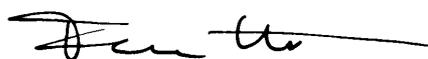
Declaration

I hereby declare that this PhD thesis contains no material which has been used to obtain any other degree at any other university or institution. To the best of my knowledge, this thesis contains no material published by other persons (except where formally quoted and referenced in the text) and has been solely written and developed by myself. The core theme of the thesis, including all ideas and developments (codes and databases, except for existing software which has been used), as well as the writing up of the text was my own responsibility.

This thesis includes three original papers published in peer-reviewed international journals and one submitted publication. The inclusion of co-authors in these four scientific papers reflects the fact that the work was a result of a research collaboration and thus acknowledges this team-based research. The co-authors have mainly contributed through the provision of data, but also by performing (not designing) analyses and through mentoring and guidance.

My contribution to the published papers involved the following:

Thesis chapter	Publication title	Nature and extent (approx. %) of my contribution
2	<i>Sphaerosyllis levantina</i> sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in <i>Sphaerosyllis hystrix</i> Claparède, 1863 (<i>published</i>)	Idea and concept: 100% Sorting and identification of the specimens: 100% Morphometric measurements: 50% Taxonomic research and description of the taxon: 90% Morphometric analyses: 80% Illustrations: 80% Paper writing: 95% Data management: 100%
3	Micro-computed tomography: Introducing new dimensions to taxonomy (<i>published</i>)	Idea and concept: 100% Scanning of specimens: 100% Genetic analyses: 10% Illustrations: 60% Analyses: 80% Paper writing: 95% Data management: 100%
4	<i>Polytraits</i> : A database on biological traits of marine polychaetes (<i>published</i>)	Idea and concept: 100% Development of database and code: 95% Data collection: 50% Paper writing: 90% Data management: 100%
5	Comparison of structural and functional stability of polychaete assemblages in coastal lagoons (<i>submitted</i>)	Idea and concept: 100% Data collection: 30% Analyses: 90% Paper writing: 90% Data management: 100%



Sarah Faulwetter

Heraklion, 05.05.2014

Abstract

Biodiversity research is currently undergoing a silent but fast revolution. The massive availability of digital biodiversity information – facilitated by the widespread use of the internet and the every-day use of personal computers in science – paves the way for unprecedented analyses and discoveries. This new data-centric discipline relies more than ever on producing, sharing and reusing digital data, yet, sound data management and data publication are still not a part of the normal scientific workflow. A variety of factors are responsible for the reluctance of scientists to make their data publicly available for reuse, but the lack of technical knowledge, coupled with the time-consuming process of converting the data into a standardised format allowing to be incorporated into appropriate databases, might be the main obstacle. Therefore, new tools, workflows and incentives which facilitate the scientist's daily work and provide acknowledgement of this work are needed. This can be achieved by providing an intuitive working environment which handles all technical aspects of data formatting and organisation, and by ensuring that researchers receive academic credit for the publication of their data.

This thesis consists of four research papers answering specific scientific questions in the field of taxonomy and ecology. The work in each paper is supported by using existing and newly developed tools and methods from the field of biodiversity informatics and cybertaxonomy, demonstrating the value of appropriate data management and publication for cutting-edge research outputs. The data underpinning the research performed in this thesis is managed and published through a virtual research environment, a Scratchpad (<http://polychaetes.lifewatchgreece.eu>), to which the four research papers forming the main body of this thesis are linked. The papers presented here demonstrate an approach to modern biodiversity research by using existing and newly created tools and by linking, sharing and mobilising data, always with the aim of making the underlying data and results reusable in a wider context. Chapter 2 provides the description of a new species, which is prepared and published through an integrated online writing tool in a virtual research environment. Chapter 3 proposes a new data type, three-dimensional digital morphological and anatomical data, and the concept of virtual type material, so-called “cybertypes”, to transform conventional taxonomy into a cyberscience. This chapter identifies important research directions for the biodiversity informatics and museum communities towards the creation of virtual collections

and the curation of potential cybertypes. Chapter 4 is the first data paper to describe a semantically annotated database of species life-cycle trait information (<http://polytraits.lifewatchgreece.eu>) and to make use of the Encyclopedia of Life's TraitBank as a repository for the data. Chapter 5 then makes use of this trait information in an ecological study, demonstrating the value of such information towards gaining new ecological insights, providing a potential approach to predicting human-induced changes to ecosystems, and underlining the importance of collecting and sharing traits data at a global scale.

Finally, this thesis identifies a number of challenges for the biodiversity informatics community which need to be addressed in the near future in order to make best use of the existing and newly created data.

Keywords: biodiversity informatics, cybertaxonomy, biological traits, data management, Polychaeta

Περίληψη

Το ερευνητικό πεδίο της βιοποικιλότητας διανύει μια περίοδο σιωπηλής αλλά ταχείας επανάστασης. Η διαθεσιμότητα μεγάλου όγκου ψηφιακής πληροφορίας βιοποικιλότητας, η ύπαρξη διαδικτυακών συνδέσεων υψηλής ταχύτητας και η χρήση υπολογιστών στην καθημερινότητα της έρευνας διανοίγουν μια οδό για καινοτόμες αναλύσεις και ανακαλύψεις χωρίς προηγούμενο. Αυτή η νέα «δεδομένο-κεντρική» κατεύθυνση της έρευνας βασίζεται όλο και περισσότερο στη παραγωγή, δημοσίευση και επαναχρησιμοποίηση ψηφιακών δεδομένων. Εντούτοις, η σωστή διαχείριση και δημοσίευση δεδομένων δεν αποτελούν ακόμα μέρος της καθημερινής επιστημονικής ενασχόλησης. Υπάρχουν συγκεκριμένοι παράγοντες υπεύθυνοι για την απροθυμία των επιστημόνων να δημοσιεύουν τα δεδομένα τους με σκοπό την επαναχρησιμοποίησή τους από την επιστημονική κοινότητα. Ο σημαντικότερος όλων φαίνεται να είναι η έλλειψη τεχνικών γνώσεων και χρόνου — απαραίτητες προϋποθέσεις για την τυποποίηση των δεδομένων, ώστε να μπορούν να ενσωματώνονται σε κατάλληλες βάσεις δεδομένων. Για την υποστήριξη του ερευνητή στη διαδικασία αυτή χρειάζονται νέα εργαλεία, ροή εργασίας και κίνητρα, όπως η παροχή εύχρηστου εικονικού περιβάλλοντος εργασίας το οποίο να αναλαμβάνει όλα τα τεχνικά βήματα της μετατροπής και οργάνωσης των δεδομένων, καθώς και ένας μηχανισμός διαπίστευσης του ερευνητή υπό τη μορφή επιστημονικών αναφορών όταν επαναχρησιμοποιούνται τα δεδομένα του.

Η παρούσα εργασία αποτελείται από τέσσερις επιστημονικές δημοσιεύσεις οι οποίες απαντούν σε ειδικά ερωτήματα στα πεδία της ταξινομικής και της οικολογίας. Η έρευνα σε κάθε δημοσίευση υποστηρίζεται από υπάρχοντα και νέα εργαλεία και μεθόδους προερχόμενα από τα πεδία της πληροφορικής της βιοποικιλότητας και της κυβερνοταξινομικής, αποδεικνύοντας την αξία της σωστής διαχείρισης και δημοσίευσης δεδομένων για ερευνητικά αποτελέσματα σε τομείς αιχμής.

Η διαχείριση και δημοσίευση των δεδομένων στα οποία βασίζονται οι έρευνες αυτές έγινε με τη βοήθεια ενός εικονικού περιβάλλοντος εργασίας, τα Scratchpads (<http://polychaetes.lifewatchgreece.eu>). Το περιβάλλον αυτό συνδέει τις τέσσερις επιστημονικές εργασίες από της οποίες αποτελείται ο πυρήνας της παρούσας διατριβής. Αυτές οι εργασίες παρουσιάζουν μια σύγχρονη προσέγγιση στη έρευνα της βιοποικιλότητας, χρησιμοποιώντας υπάρχοντα εργαλεία και μεθόδους, συνδέοντας και κινητοποιώντας επιστημονικά δεδομένα με σκοπό την επαναχρησιμοποίησή τους σε ένα ευρύτερο πλαίσιο.

Το Κεφάλαιο 1 παρέχει μια εισαγωγή στο πεδίο της πληροφορικής της βιοποικιλότητας και της

κυβερνοταξινομικής και περιγράφει τη δομή της διατριβής αυτής.

Το Κεφάλαιο 2 παρουσιάζει τη περιγραφή ενός νέου είδους Πολυχαίτου χρησιμοποιώντας σύγχρονα πρότυπα και εργαλεία της πληροφορικής της βιοποικιλότητας. Η περιγραφή του είδους εκπονήθηκε αποκλειστικά στο Scratchpads, συγκεντρώνοντας όλες της απαραίτητες πληροφορίες και υλικό (π.χ. βιβλιογραφία, φωτογραφίες, δεδομένα κατανομής) μέσα σε αυτό το δομημένο διαδικτυακό περιβάλλον. Η περιγραφή του είδους αποθηκεύτηκε ως ειδικός τύπος δεδομένων («Taxon Description»). Με τη βοήθεια ενός ενσωματωμένου στο Scratchpads περιβάλλοντος για κοινή και διαδικτυακή ολοκλήρωση του άρθρου (μονάδα δημοσίευσης, Publication Module), τα χωριστά κομμάτια πληροφορίας σχετικά με το νέο είδος συναθροίστηκαν σε χειρόγραφο το οποίο στάλθηκε αυτόματα στο περιοδικό ως έγγραφο σήμανσης το οποίο μπορεί να υποστεί επεξεργασία από υπολογιστές.

Στο Κεφάλαιο 3 προτείνεται ένα νέο είδος δεδομένων: τρισδιάστατα ψηφιακά μορφολογικά και ανατομικά δεδομένα, και προτείνεται επιπλέον η ιδέα του ψηφιακού υλικού-τύπου («cybertypes») ως θεμελιώδες βήμα στη μετατροπή της κλασικής ταξινομικής σε μια «κυβερνοεπιστήμη» (cyberscience). Μια σχετικά νέα τεχνική απεικόνισης, η υπολογιστική μικροτομογραφία ακτίνων X (ή απλά μικροτομογραφία ή micro-CT), χρησιμοποιήθηκε για την δημιουργία αυτών των ψηφιακών μορφολογικών και ανατομικών δεδομένων. Η μικροτομογραφία είναι μια νέα τεχνική παρόμοια με την ιατρική τομογραφία αλλά εφαρμόζεται σε κλίμακες μεγέθους χιλιοστού ή εκατοστού και παράγει εικόνες υψηλότερης διακριτικής ικανότητας. Η τεχνική είναι ιδανική για την απεικόνιση βιολογικών δειγμάτων επειδή επιτρέπει τη μελέτη μορφολογικών και ανατομικών χαρακτηριστικών χωρίς να καταστρέφει το πολύτιμο υλικό. Επιπλέον, φαίνεται ότι δεν αποδομεί τις αλληλουχίες νουκλεϊκού οξέος οι οποίες επιτρέπουν να γίνει μοριακή ανάλυση ενός δείγματος ώστε να προσδιοριστεί η γενετική σύσταση ενός είδους. Επομένως, η τεχνική έχει μεγάλες δυνατότητες για τη δημιουργία ψηφιακών ειδώλων μουσειακού υλικού, (και ιδιαίτερα του υλικού-τύπου) τα οποία θα μπορούσαν να χρησιμοποιούνται ως «cybertypes» και για τη δημιουργία εικονικών συλλογών. Σκοπός της μελέτης ήταν α) να καθοριστούν οι απαιτήσεις για τα εν λόγω «cybertypes», β) η αξιολόγηση της τεχνικής σχετικά με τις δυνατότητες και τους περιορισμούς της για τη δημιουργία των «cybertypes», γ) ο προσδιορισμός σημαντικών κατευθύνσεων στο πεδίο της πληροφορικής της βιοποικιλότητας και της μουσειολογίας σχετικά με τη δημιουργία εικονικών συλλογών και τη διαχείριση του ψηφιακού υλικού-τύπου και δ) η επίδειξη μιας νέας μορφής ταξινομικής δημοσίευσης η οποία ενσωματώνει διαδραστικά τρισδιάστατα μοντέλα στα ηλεκτρονικά κείμενα.

Το Κεφάλαιο 4 είναι η πρώτη εργασία σε μορφή «δημοσίευσης δεδομένων» (data paper) η οποία περιγράφει τη δημιουργία μιας σημασιολογικά εμπλουτισμένης βάσης δεδομένων με

πληροφορίες για βιολογικά χαρακτηριστικά των ειδών, το *polytraits* (<http://polytraits.lifewatchgreece.eu>). Σκοπός της μελέτης ήταν η ψηφιοποίηση και η διάδοση πληροφορίας σχετικά με βιολογικά χαρακτηριστικά των Πολυχαίτων. Δεδομένα για βιολογικά χαρακτηριστικά (τα οποία περιγράφουν, π.χ. τον κύκλο ζωής, τη συμπεριφορά ή τη μορφολογία των οργανισμών) συνήθως δεν δημοσιεύονται από την επιστημονική κοινότητα, ή σε μορφές οι οποίες δεν επεξεργάζονται από υπολογιστές (π.χ. σε παραρτήματα επιστημονικών δημοσιεύσεων). Στη παρούσα εργασία αναπτύχθηκε η βάση δεδομένων και συλλέχθηκαν και εισήχθηκαν δεδομένα από τη βιβλιογραφία. Τα δεδομένα διατίθενται σε διάφορες μορφές, ώστε να μπορούν να χρησιμοποιηθούν από ανθρώπους, αλλά και σε μορφές οι οποίες αναγνωρίζονται και μπορούν να ερμηνευτούν από υπολογιστές. Τελικά, τα δεδομένα υποβλήθηκαν και στην «TraitBank» της Εγκυκλοπαίδειας της Ζωής (EOL), ένα νέο αποθετήριο για πληροφορίες σχετικές για βιολογικά χαρακτηριστικά των ειδών.

Στο Κεφάλαιο 5 τα δεδομένα από το *polytraits* αναλύονται στο πλαίσιο μιας οικολογικής υπόθεσης. Η εργασία αυτή αποδεικνύει την αξία αυτού του είδους πληροφορίας για την διερεύνηση ειδικών οικολογικών ζητημάτων, οι οποίες παρέχουν μία εκτίμηση πιθανών αλλαγών της λειτουργικότητας του οικοσυστήματος και υποστηρίζουν τη διαχείριση οικοσυστημάτων. Συγκεκριμένα, μελετήθηκε ο τρόπος με τον οποίο αλλάζουν τα πρότυπα κατανομής των βιοκοινωνιών των Πολυχαίτων σε 6 λιμνοθάλασσες της νοτιοανατολικής Ευρώπης καθώς και τα πρότυπα κατανομής των βιολογικών χαρακτηριστικών σε υποθετικά σενάρια απώλειας ειδών. Η εργασία αυτή αποδεικνύει την αξία αυτού του είδους πληροφορίας για την διερεύνηση ειδικών οικολογικών ζητημάτων, οι οποίες παρέχουν μία εκτίμηση πιθανών αλλαγών της λειτουργικότητας του οικοσυστήματος και διαθέτουν επιστημονική πληροφορία για τη διαχείριση οικοσυστημάτων. Η εργασία τονίζει τη σημαντικότητα της συλλογής και δημοσίευσης πληροφορίας για βιολογικά χαρακτηριστικά σε παγκόσμια κλίμακα.

Το Κεφάλαιο 6 παρέχει μια συζήτηση των ερευνών οι οποίες διεξήχθησαν στο πλαίσιο της παρούσας διατριβής και τις εντάσσει σε ένα ευρύτερο πλαίσιο. Το ύστατο κεφάλαιο της παρούσας διατριβής είναι ο προσδιορισμός σειράς από σημαντικών ερευνητικών προκλήσεων στο πεδίο της πληροφορικής της βιοποικιλότητας οι οποίες πρέπει να αντιμετωπιστούν στο άμεσο μέλλον με σκοπό τη σωστή διαχείριση και χρήση πολύτιμων δεδομένων και πληροφοριών.

Λέξεις κλειδιά: πληροφορική της βιοποικιλότητας, κυβερνοταξινόμηση, βιολογικά χαρακτηριστικά, διαχείριση δεδομένων, Πολύχαιτοι

Acknowledgements

This thesis would not have been possible without the outstanding support and encouragement I received from a number of persons. Special thanks go to my research adviser, Dr. Christos Arvanitidis for his enthusiasm and encouragement to pursue my work at the interface between two disciplines and for all the support he has provided by guiding my first steps as a member of the national and international research community. I would also like to express my gratitude to my two academic advisors, Prof. Artemis Nicolaidou and Prof. Anastasios Legakis for supervising and encouraging my research and assisting me with all the requirements of the academic system. Many thanks go also to my examination committee, Dr. Martin Dörr, Prof. Chariton Chintiroglou, Prof. Theodoros Kevrekidis and Prof. Alberto Basset.

I furthermore thank all my colleagues from the Laboratory of Biodiversity in the Hellenic Centre for Marine Research who have always had an open ear or a solution for all the seemingly insurmountable problems I have encountered during the course of this research, and, most importantly for their generosity in the multiple cases I needed a helping hand to accomplish a task.

I also received outstanding support from a number of colleagues from Greece and abroad – each of these persons is acknowledged for their contributions in the respective chapter of this thesis. However, I would like to express my special thanks to Dr. Lyubomir Penev, CEO of Pensoft Publishers, for always being open to new ideas concerning the publication of research papers in unconventional formats and for promoting this work to the international research community and beyond. I also thank Dr. Vincent Smith and Dr. Dave Roberts of the Natural History Museum in London for their creativity and enthusiasm concerning cybertaxonomic research which has been a continuous source of inspiration for this work.

Preface

The research for this thesis has been conducted in the “trading zone” between two different disciplines, biology and informatics. The use of specific terms and expressions from both fields in the text could not be avoided, even though they may not be easily understandable by the casual reader. To facilitate understanding, an extensive glossary of terms and abbreviations is provided at the end of this thesis which can be consulted for additional explanations which are not provided in the published text.

Technology is a fast-evolving field, and over the course of this thesis certain aspects which were state-of-the-art at the time of writing are now considered outdated. The research environment used in this thesis is such an example: the initial software version, Scratchpads 1.0 has now been replaced by a much more sophisticated version, Scratchpads 2.0. Chapters 2 and 3 of this thesis are still based on Scratchpads 1.0, but during the work on Chapters 4 and 5 it became obvious that an upgrade of the system was needed to accommodate the needs for certain developments. The process of upgrading is still ongoing, and there is still some content in the old version of the Scratchpads which has not yet been migrated to the new site. Furthermore, due to “political” decisions, the web address of the new Scratchpads installation had to be changed as well: the old address, <http://polychaetes.marbigen.org> will continue to be accessible but once the data migration process is complete this address will redirect the user to the new installation of the software, <http://polychaetes.lifewatchgreece.eu>. For the purposes of the evaluation of this thesis, both installations are still accessible and run in parallel, allowing the examiners to assess the developments described in the first two publications of this thesis.

Chapter 1

Introduction

Introduction

Introduction to Biodiversity Informatics and Cybertaxonomy

Computers have been used in biodiversity research and taxonomy already since the 1960s (Edwards and Morse 1995). Only during the last two decades, however, the digital management and the massive availability of biodiversity information have undergone a real revolution. The widespread use of the internet, fast data transfer and inexpensive computing power, as well as the use of personal computers as an integral part of scientists' daily work, have paved the way for unprecedented possibilities to use and analyse biodiversity information. As a result, completely new disciplines such as *biodiversity informatics* and *cybertaxonomy* have emerged. Biodiversity informatics is the application of informatics to biodiversity information and includes all steps from digital data management to analysis and dissemination of taxonomic, biogeographic and ecological information. *Biodiversity informatics* is not synonymous with *bioinformatics*: the former is applied to information pertaining to the organism level and above, whereas the traditionally latter concerns information originating from molecular biology (Soberón and Peterson 2004). The term *biodiversity informatics* was coined in the 1990s¹ and had gained entry into the scientific literature by the end of the decade (e.g. Schalk 1998, Bisby 2000). The term *cybertaxonomy* is even more recent in origin, having originated in the mid-2000s. It can be regarded as a sub-discipline of biodiversity informatics, describing all informatics-related work practices that are used by taxonomists *sensu lato* to support and accelerate their work (Wheeler 2010).

One of the first milestones in the community efforts in the field of biodiversity informatics was the foundation of the Global Biodiversity Information Facility (GBIF)² in 2001 – a global initiative to assemble biodiversity information and make it freely accessible over the internet. Complementary initiatives soon followed: a) the Ocean Biogeographic Information System³ started to assemble, integrate and disseminate marine species distribution data; b) online taxonomic databases appeared in an attempt to provide a standardised taxonomy to the community (e.g. the Catalogue of Life / Species 2000 initiative (COL)⁴, the World Register of

1 <https://web.archive.org/web/20130511091435/http://www.bgbm.org/BioDivInf/TheTerm.htm>, last accessed on 2013-11-30

2 <http://www.gbif.org>, last accessed on 2013-11-30

3 <http://www.iobis.org>, last accessed on 2013-11-30

4 <http://www.catalogueoflife.org>, last accessed on 2013-11-30

Marine Species (WoRMS)⁵, the Global Names Architecture (GNA)⁶, Index Fungorum (IF)⁷, the International Plant Names Index (IPNI)⁸, ZooBank⁹, or the Index of Organism Names (ION)¹⁰; and c) species information systems (e.g. Fishbase¹¹, Algaebase¹², Wikispecies¹³, the Encyclopedia of Life¹⁴ or the Atlas of Living Australia¹⁵) were developed to provide additional data on species (images, literature references, ecological information, distribution maps, species descriptions, etc.) and make these data accessible through centralised web portals. A large part of the scientific literature is nowadays also available electronically and can be retrieved over the internet, and legacy biodiversity literature is increasingly being digitised and made available by the Biodiversity Heritage Library¹⁶, Google Books¹⁷ or the Internet Archive¹⁸, to name just a few initiatives.

However, despite the fast pace that the discipline itself is setting, it is still facing several unsolved problems, as new developments introduce new challenges (Hardisty et al. 2013). One of these issues is that the basic community – the data producers – have not yet fully adapted to this new digital world, and this is reflected in working practices (e.g. Chavan and Ingwersen 2009, Tenopir et al. 2011). Most biodiversity data are still produced in analogue formats – or in digital formats that remain on scientist's desktops or local servers or are “locked up” in formats that cannot readily be exploited by computers. However, digital data that are not connected through a standard format to the data network are “invisible” to the rest of the world and thus little more useful to the global community than a hand-written lab book. This underdeveloped data sharing culture creates a huge problem to the very foundations of the discipline. In contrast to the opinions and interpretations communicated through scientific publications – which are end-products of the scientific process –, primary biodiversity and ecological data can be used and re-used in various contexts and thus form the basis for testing new hypotheses and the creation of new knowledge. The value of these data grow over time, since they document the

5 <http://www.marinespecies.org>, last accessed on 2014-01-20

6 <http://www.globalnames.org>, last accessed on 2014-01-20

7 <http://www.indexfungorum.org>, last accessed on 2013-11-30

8 <http://www.ipni.org>, last accessed on 2013-11-30

9 <http://zoobank.org>, last accessed on 2013-11-30

10 <http://www.organismnames.com>, last accessed on 2013-11-30

11 <http://www.fishbase.org>, last accessed on 2013-11-30

12 <http://www.algaebase.org>, last accessed on 2013-11-30

13 <http://species.wikimedia.org>, last accessed on 2013-11-30

14 <http://www.eol.org>, last accessed on 2014-01-20

15 <http://www.ala.org.au>, last accessed on 2014-01-20

16 <http://www.biodiversitylibrary.org>, last accessed on 2013-11-30

17 <http://books.google.com>, last accessed on 2014-01-15

18 <https://archive.org>, last accessed on 2013-11-30

state of a dynamic system at a specific point in space and time and are thus unreproducible (Costello 2009). This fact has been recognised by the scientific community (Nature Editorial 2005, Enke et al. 2012, Costello et al. 2013, Hardisty et al. 2013) and might be attributed to a variety of reasons, two of which are addressed in this thesis:

- a) *Lack of incentives for data sharing.* Scientists, especially in biodiversity sciences, are reluctant to share their data which often require years and much effort to assemble. The perception seems to hold that by releasing the data, others will take undue credit for potential future analyses based on these data, implying a loss of ownership and ultimately control over the dataset (Enke et al. 2012). Furthermore, scientists are rewarded for their published work through citations by peers – a metric of acknowledgement for their contribution to science, This metric is nowadays also used as a proxy to evaluate the respective researcher's achievements and further career advancements. For the publication and re-use of datasets there are no equivalent measures yet, thus a primary motivation factor for the publication of data – academic credit – is missing (Costello 2009). In addition, the scarcity of legal requirements or guidelines by journals to make data available to the public, adds to the disinterest of researchers working in the field of biodiversity to make their data available.
- b) *Lack of expertise to appropriately publish the data.* Data publication requires that the data are converted into formats that are compatible with recommended standards (e.g. the Darwin Core standard, Wieczorek et al. (2012)), annotated with metadata, quality-controlled and published through appropriate channels (e.g. biogeographic databases such as OBIS or GBIF). Since this process requires additional time and expertise, it is often neglected, and in combination with the lack of incentives, it results in many valuable primary data being unavailable to the community (Smith et al. 2011, Costello and Wieczorek 2013). What is needed, therefore, are new tools which facilitate and speed up data publication without requiring technical knowledge, and at the same time allow the researcher to receive academic credit for the data. Such tools need to be as easy to use as any other common desktop software and perform all required formatting, standardisation and management of the data without requiring specific skills by the user.

An additional gap in the current biodiversity informatics landscape is the absence of certain types of data from the growing network of information. Almost no digital information is available on:

- a) species attributes and species interactions, a type of information which is needed to answer questions about ecological processes and to create general ecosystem models that can ultimately be used for predictive modelling of the biosphere (Harfoot and Roberts 2014)
- b) digital, quantitative morphological information provided by the myriads of type specimens stored in museums and collections around the globe (Smith and Blagoderov 2012). These specimens are the foundation for taxonomic research and are thus essential for the definition of the working units for biodiversity research: the species. In the digital age, morphological information coded as text in taxonomic descriptions is not sufficient to perform quantitative analyses across the tree of life, but the actual morphology of the specimen needs to be digitised in the form of images, three-dimensional models or other quantitative data formats which can then be subjected to powerful analyses.

These two types of data urgently need to be made accessible digitally, but in contrast to taxonomic and biogeographic data, no infrastructure or standards exist so far which would facilitate a structured digitisation and publication of this information.

This thesis contributes to the two main issues defined above through:

- a) the application and development of new methods and tools to facilitate and accelerate taxonomic and ecological research
- b) the creation and publication of two hitherto neglected data types: morphological and anatomical data in the form of virtual specimens, and species life-cycle data.

In both cases the limitations of the current biodiversity cyberinfrastructure become obvious. Those parts of the cyberinfrastructure which have the potential to pave the way for new initiatives and research directions are discussed and highlighted with specific examples.

Structure of the thesis

The main body of this thesis comprises four chapters (Chapter 2 – Chapter 5) which have been published in or submitted to international peer-reviewed journals in the form of independent papers. Chapters 2 and 3 focus on different aspects in cybertaxonomy: the description of a new species through a new, innovative workflow and publication format (Chapter 2) and the introduction of virtual, three-dimensional specimens to be used in taxonomic research (Chapter

3). Chapters 4 and 5 focus on species' life-cycle traits, forming a thematic entity: chapter 4 is a data paper – a publishing format introduced only recently, formally describing a dataset – and chapter 5 demonstrates the application of this information in ecosystem research. Chapter 6 provides a discussion on this thesis and integrates the results into a wider context.

Common to all papers is their data-centric approach: all research performed in this thesis is based on modern information management methods and tools, which, in turn, serve the purpose of these papers to test specific scientific hypotheses. To manage all information, existing web-based software, databases and standards are used and integrated with own developments (Figure 1). Since the research papers do not focus on the underlying data management infrastructure in detail but on their respective scientific hypotheses, a short outline of each chapter in the context of this thesis, as well as of the virtual research environment are given below.

Virtual research environment and data architecture / data flow

Central to the data and information underlying this research is a Scratchpad¹⁹ (Smith et al. 2011). Scratchpads²⁰ are virtual research environments – a web-based software which facilitates the organisation and publication of biodiversity data. Data are organised into different types of information – e.g. images, videos, specimen information, literature, species descriptions, occurrences, etc. – and are organised around a biological classification. Each piece of information can be tagged with a taxon name, and thus the information can be browsed either by navigating the biological classification or by searching for the taxon name. All information pertaining to a taxon is then displayed on so-called “taxon pages”. The system is easy to use and no special technical knowledge is required. Its communal design allows groups of researchers to use the system simultaneously, to collaboratively work on a project and to share data, either publicly or privately within virtual research groups. Data are formatted according to the Darwin Core standard – the community data standard primarily used by the biodiversity informatics community – and are ready to be shared and integrated with other data (e.g. submitted to or harvested by large global data aggregators such as EOL or GBIF) without requiring further technical knowledge by the user. In return, by using Darwin Core, data from existing databases and projects can easily be imported into Scratchpads and displayed on species pages (Baker et al. 2014).

19 <http://polychaetes.lifewatchgreece.org>

20 <http://scratchpads.eu>, last accessed on 2014-01-20

However, the Scratchpads architecture is not an ideal solution for storing and disseminating all types of data. Certain data types require an infrastructure which is not provided by the database structure underlying the software. Therefore, additional software has been developed in the course of this thesis, and certain data have been archived in external repositories such as DataDryad²¹ or figshare²². Specifically, the management of species life-cycle information demanded a much more complex data structure and data entry interface than the one available by the Scratchpads²³. Therefore, a dedicated database with a user interface was developed to collect and store these data: the *polytraits* database²⁴. The data from this separate database can however be accessed through the central Scratchpads interface: when a taxon page is opened on the Scratchpads, the system sends a request to the *polytraits* database, retrieves any available trait information for the requested taxon and displays this information (Figure 1; for a screenshot of the data integration see Chapter 4, Figure 1).

Scratchpads can, therefore, not only be used to organise and publish information uploaded directly to the system, but also to import or display information from remote sources (Smith et al. 2009). Existing Scratchpad modules already allow the user to integrate information from EOL, BHL, Google Scholar or GBIF, but this work demonstrates that with programming skills, third-party content can be easily integrated by writing new modules for the software.

Outline of chapters

Chapter 1

This first chapter provides an introduction to the field of biodiversity informatics and cybertaxonomy and describes the structure of the thesis.

Chapter 2

The aim of this work was the description of a new polychaete species (bristle worms, Annelida: Polychaeta) using modern biodiversity informatics standards and tools, thus exploring and exemplifying methods which have the potential to support, enhance and accelerate the work of taxonomists. The description of the new species has been prepared entirely in the Polychaetes Scratchpad by assembling all information (e.g. media, bibliography, distribution data)

21 <https://datadryad.org>, last accessed on 2013-11-30

22 <http://www.figshare.com>, last accessed on 2014-01-20

23 It should be noted that since this thesis was started the Scratchpads software has constantly been improved, and in its current version the integration of traits information would not be as complicated as it was during an earlier version of this software.

24 <http://polytraits.lifewatchgreece.eu>

pertaining to the new taxon (and related species) in a structured online working environment. The description of the taxon was stored as a special media type (Taxon Description). The built-in *Publication Module* was subsequently used to assemble the various bits and pieces into a coherent manuscript, with the option to add additional sections. The completed manuscript was finally sent to the journal (Zookeys) with a single click. At this stage, an XML representation of the manuscript is created (according to the TaxPub extension²⁵), allowing the data to be further processed by computers. A detailed overview of the steps leading from the manuscript preparation in the Scratchpads to the final publication can be found in Penev et al. (2011)²⁶.

Chapter 3

This chapter explores the potential of a modern, fully computerized, three-dimensional imaging technique – micro-computed tomography (also called *micro-CT* or *X-ray microtomography*) – to create digital morphological and anatomical data to support systematic and taxonomic studies. Micro-CT scanning, which is similar to computer tomography used in medicine, can be used to create three-dimensional virtual representations of the internal and external features of the imaged object in a sub-micron resolution. This makes the technique extremely interesting for imaging biological specimens, primarily because morphological and anatomical features can be studied without having to dissect the specimen, and the process does not appear to destroy the nucleic acid sequences which are used for the genetic identification of the specimen. The technique has a strong potential to create virtual representations of museum specimens, especially those of the type material, which could be used as “cybertypes”, and to assemble these representations into virtual museum collections. The aim of this study was a) to define requirements for such cybertypes, b) to evaluate the imaging technique and its potential and limitations for creating cybertypes, c) to identify challenges for the biodiversity informatics and cybertaxonomy community concerning the creation of virtual galleries and the curation of cybertypes, and d) to demonstrate new formats of publishing taxonomic works by embedding interactive, three-dimensional models into scientific publications. The videos and interactive three-dimensional models of the scanned specimens are embedded into the PDF version of the manuscript and can be played and interactively manipulated if viewed with Acrobat Reader (Version 8 or higher). All derivative media included in the publication as well as supporting

25 <http://sourceforge.net/projects/taxpub>, last accessed on 2014-01-22

26 Also included in Annex I of this thesis

material (surface models, image files, videos) were published through the Polychaete Scratchpads, and the full volumetric datasets were archived in the Dryad Data Repository (doi:10.5061/dryad.84m54).

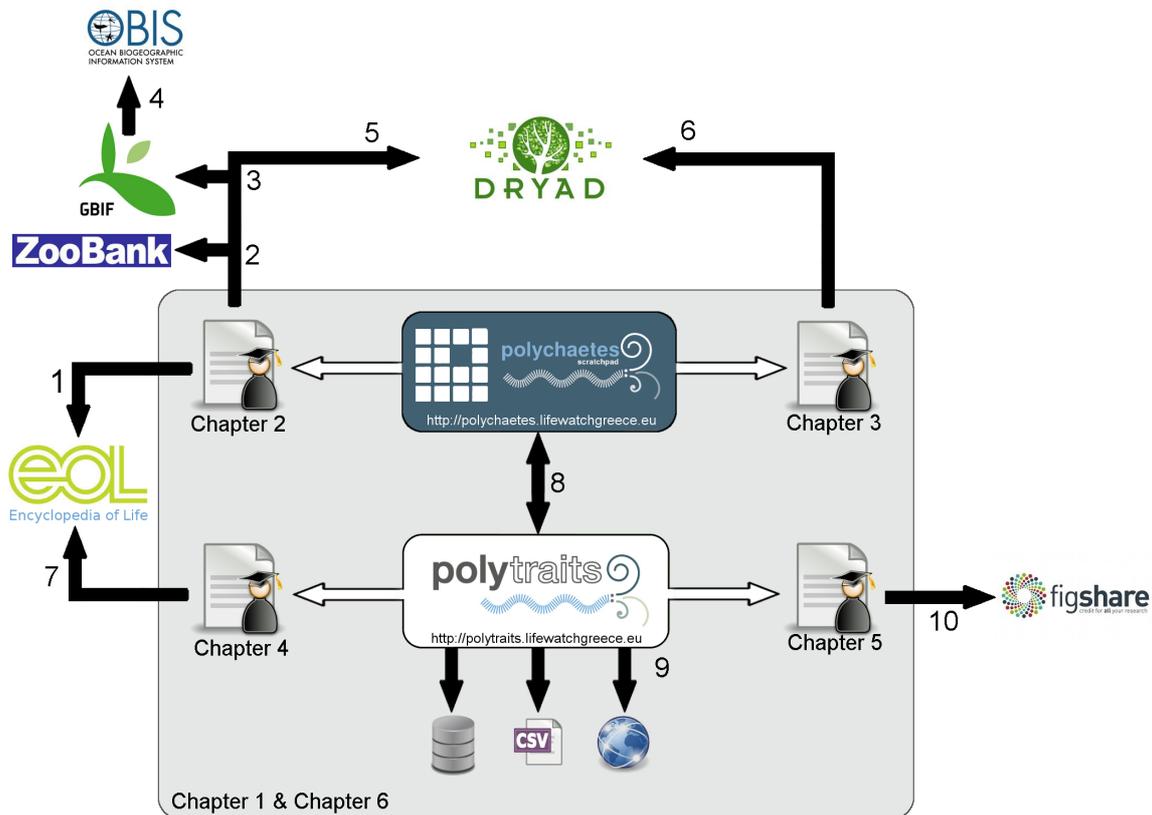


Figure 1: Schematic representation of the components and the data flow in this thesis. Black arrows: data flow, white arrows: data analysis and interpretation. **1:** Description of the new species and images are deposited in EOL. **2:** Nomenclatural act is registered in Zoobank. **3:** Distribution data are sent to GBIF and from there to **4:** OBIS (via an IPT installation). **5:** Darwin Core-formatted file of distribution data is also deposited in Dryad. **6:** Volumetric datasets are deposited in Dryad. **7:** Traits are deposited in EOL's TraitBank. **8:** Data from the *polytraits* database are displayed in Scratchpads via an API. **9:** Data from the *polytraits* database are provided in various download formats. **10:** Data and scripts from Chapter 5 are deposited in figshare.

Chapter 4

The aim of this study was to digitise and disseminate information on the biological traits of polychaetes. Trait data, which describe e.g. morphological, behavioural and life-cycle characteristics, are either not published at all by research communities, or only in forms which cannot readily be processed by computers, such as supplementary documents to publications.

The absence of this information from the global biodiversity data network might be attributed to the absence of trait data repositories and data standards for this type of information. During this study, *polytraits*, a database on biological traits of polychaetes was developed and populated with information extracted from the literature. The data are made available in various formats so that they can not only be read by humans, but they also are formatted according to standards which can be recognised and properly interpreted by computers. The *polytraits* database not only stores the data but also includes: a) a public interface which provides a documentation of the data structure and the contents; b) a variety of download options; c) a password-protected data entry interface (see Appendix II for screenshots and documentation). The actual contents of the database can be viewed through the Polychaetes Scratchpad. Currently, the database contains almost 20,000 records on morphological, behavioural and reproductive characteristics of more than 1,000 polychaete species, all referenced by literature sources and accompanied by the respective text passage from the literature (for reasons of quality control, future re-organisation of the data and their re-usability in different contexts). In collaboration with the Encyclopedia of Life (EOL, a mega-project on biodiversity informatics, based in the US – Marine Biology Laboratory, Woods Hole and Smithsonian Institution, Washington), the *polytraits* data are also integrated into EOL's TraitBank²⁷, a recently launched initiative aiming at becoming a global digital repository for information on species' life-cycles and interactions²⁸.

Chapter 5

This study makes use of the trait information assembled through the *polytraits* database and demonstrates how this type of information provides new ecological insights. Specifically, the effects of the loss of species on the taxonomic and functional diversity in six coastal Mediterranean lagoons is simulated. The study, therefore, provides an initial assessment of the “what if” type of questions, through the comparison of the information patterns produced by the different sources of information: the traditionally used community structure data (i.e. species composition and abundance) and biological attributes data, which reflect the role of the species in the ecosystem. The raw data and scripts used in this study are freely available through figshare (doi:10.6084/m9.figshare.909347).

27 <http://eol.org/traitbank>, last accessed on 2014-01-22

28 <http://blog.eol.org/post/74751984298/announcing-traitbank>, last accessed on 2014-01-27

Chapter 6

The last chapter discusses the research conducted in this thesis in the context of current trends, activities and research directions in the field of biodiversity informatics and cybertaxonomy and provides concluding remarks.

References

- Baker E, Rycroft S, Smith VS (2014) Linking multiple biodiversity informatics platforms with Darwin Core Archives. *Biodiversity Data Journal* 2: e1039. doi:10.3897/BDJ.2.e1039
- Bisby FA (2000) The quiet revolution: biodiversity informatics and the internet. *Science* 289: 2309–12.
- Chavan VS, Ingwersen P (2009) Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. *BMC Bioinformatics* 10 Suppl 1: S2. doi:10.1186/1471-2105-10-S14-S2
- Costello MJ (2009) Motivating Online Publication of Data. *BioScience* 59: 418–427. doi:10.1525/bio.2009.59.5.9
- Costello MJ, Michener WK, Gahegan M, Zhang Z-Q, Bourne PE (2013) Biodiversity data should be published, cited, and peer reviewed. *Trends in ecology & evolution* 28: 454–61. doi:10.1016/j.tree.2013.05.002
- Costello MJ, Wiczorek J (2013) Best practice for biodiversity data management and publication. *Biological Conservation*. doi:10.1016/j.biocon.2013.10.018
- Edwards M, Morse DR (1995) The potential for computer-aided identification in biodiversity research. *Trends in Ecology & Evolution* 10: 153–158.
- Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B (2012) The user's view on biodiversity data sharing – Investigating facts of acceptance and requirements to realize a sustainable use of research data –. *Ecological Informatics* 11: 25–33. doi:10.1016/j.ecoinf.2012.03.004
- Hardisty A, Roberts D, Addink W, Aelterman B, Agosti D, *et al.* (2013) A decadal view of biodiversity informatics: challenges and priorities. *BMC ecology* 13: 16. doi:10.1186/1472-6785-13-16
- Harfoot M, Roberts D (2014) Taxonomy: Call for ecosystem modelling data. *Nature* 505: 160. doi:10.1038/505160a
- Nature Editorial (2005) Let data speak to data. *Nature* 438: 531. doi:10.1038/438531a
- Penev L, Blagoderov V, Georgiev T, Rycroft S, Scott B, Faulwetter S (2011) ViBRANT deliverable 6.1: XML markup tools and service. 10pp. Available from: <http://vbrant.eu/sites/vbrant.eu/files/D6-1-PENSOFT-XML%20mark%20up%20tool.pdf>
- Schalk P (1998) Management of marine natural resources through by biodiversity informatics. *Marine Policy* 22: 269–280. doi:10.1016/S0308-597X(98)00013-X
- Smith VS, Blagoderov V (2012) Bringing collections out of the dark. *ZooKeys* 209: 1–6. doi:10.3897/zookeys.209.3699

- Smith VS, Rycroft SD, Brake I, Scott B, Baker E, Livermore L, Blagoderov V, Roberts D (2011) Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science. *ZooKeys* 150: 53–70. doi:10.3897/zookeys.150.2193
- Smith VS, Rycroft SD, Harman KT, Scott B, Roberts D (2009) Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life. *BMC bioinformatics* 10 Suppl 1: S6. doi:10.1186/1471-2105-10-S14-S6
- Soberón J, Peterson AT (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 359: 689–98. doi:10.1098/rstb.2003.1439
- Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M (2011) Data sharing by scientists: practices and perceptions. *PloS one* 6: e21101. doi:10.1371/journal.pone.0021101
- Wheeler QD (2010) Cybertaxonomy and Ecology. *Nature Education Knowledge* 1: 6.
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard Sarkar IN (Ed). *PLoS ONE* 7: e29715. doi:10.1371/journal.pone.0029715

Chapter 2

Faulwetter S, Chatzigeorgiou G, Galil BS, Nicolaidou A, Arvanitidis C (2011) *Sphaerosyllis levantina* sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in *Sphaerosyllis hystrix* Claparède, 1863.

ZooKeys 150: 327–345.

doi: 10.3897/zookeys.150.1877

ZooKeys 150: 327–345 (2011)
doi: 10.3897/zookeys.150.1877
www.zookeys.org

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Sphaerosyllis levantina sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in *Sphaerosyllis hystrix* Claparède, 1863

Sarah Faulwetter^{1,4†}, Georgios Chatzigeorgiou^{2,4‡}, Bella S. Galil^{3§},
Artemis Nicolaidou^{1,1}, Christos Arvanitidis^{4¶}

1 Department of Zoology-Marine Biology, Faculty of Biology, National and Kapodestrian University of Athens, Panepistimiopolis, 15784, Athens, Greece **2** Department of Biology, University of Crete, 71409 Heraklion, Crete, Greece **3** National Institute of Oceanography, Israel Oceanographic & Limnological Research, POB 8030, Haifa 31080, Israel **4** Institute of Marine Biology and Genetics, Hellenic Centre for Marine Research, 71003 Heraklion, Crete, Greece

† [urn:lsid:zoobank.org:author:9BF02566-AF30-47EB-840E-DFC841B6FF84](https://doi.org/urn:lsid:zoobank.org:author:9BF02566-AF30-47EB-840E-DFC841B6FF84)

‡ [urn:lsid:zoobank.org:author:E3A716D4-9C20-4DD0-A231-EAA04168F17D](https://doi.org/urn:lsid:zoobank.org:author:E3A716D4-9C20-4DD0-A231-EAA04168F17D)

§ [urn:lsid:zoobank.org:author:06EF9833-A3C5-48FA-BBA8-1881AC51E361](https://doi.org/urn:lsid:zoobank.org:author:06EF9833-A3C5-48FA-BBA8-1881AC51E361)

| [urn:lsid:zoobank.org:author:9FA2A4EE-7E52-4111-A9B6-CDB99E7C909E](https://doi.org/urn:lsid:zoobank.org:author:9FA2A4EE-7E52-4111-A9B6-CDB99E7C909E)

¶ [urn:lsid:zoobank.org:author:737F149F-C30C-42EB-A690-5E693AD95427](https://doi.org/urn:lsid:zoobank.org:author:737F149F-C30C-42EB-A690-5E693AD95427)

Corresponding author: Sarah Faulwetter (sarifa@hcmr.gr)

Academic editor: C. Glasby | Received 3 August 2011 | Accepted 18 October 2011 | Published 28 November 2011

[urn:lsid:zoobank.org:pub:18BC77D8-3A1A-433F-AAC2-47534DF3FC48](https://doi.org/urn:lsid:zoobank.org:pub:18BC77D8-3A1A-433F-AAC2-47534DF3FC48)

Citation: Faulwetter S, Chatzigeorgiou G, Galil BS, Nicolaidou A, Arvanitidis C (2011) *Sphaerosyllis levantina* sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in *Sphaerosyllis hystrix* Claparède, 1863. In: Smith V, Penev L (Eds) e-Infrastructures for data publishing in biodiversity science. ZooKeys 150: 327–345. doi: 10.3897/zookeys.150.1877

Abstract

Examination of polychaete specimens from Haifa Bay (Israel, eastern Mediterranean Sea) revealed several individuals exhibiting morphological characteristics similar to *Sphaerosyllis hystrix* Claparède, 1863. A detailed morphometrical analysis of the Israeli specimens in comparison to specimens of *S. hystrix* and *S. boeroi* Musco, Çinar & Giangrande, 2005 supported the description of the former as a new species, *S. levantina* sp. n. Individuals of *S. hystrix* formed a very heterogeneous group with strong character variations in the analysis and the presumed cosmopolitan distribution of the species is discussed based on literature records.

Copyright Sarah Faulwetter et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Keywords

Polychaetes, Syllidae, Exogoninae, *Sphaerosyllis*, new species, Mediterranean, Cybertaxonomy, Scratchpads

Introduction

The polychaete genus *Sphaerosyllis* Claparède, 1863 (Annelida) is one of the most species-rich genera of the syllid subfamily Exogoninae. At present, ca. 48 species are considered valid within *Sphaerosyllis* after the recent split of the group into the three genera *Sphaerosyllis*, *Prosphaerosyllis* and *Erinaceusyllis* (San Martín 2005). Up to date, 18 species of the genus have been recorded from the Mediterranean Sea (Musco and Giangrande 2005), one of them described but yet unnamed (San Martín 2003), another one in the process of description (Del Pilar-Ruso and San Martín in press). In the framework of a project focusing on the soft bottom benthos of Haifa Bay (Israel, eastern Mediterranean Sea), a number of individuals of the genus *Sphaerosyllis* were found to exhibit morphological features which did not entirely correspond to any description of known *Sphaerosyllis* species, namely falcigers with a strong serration and with a subdistal spine present in all chaetigers. A subdistal spine on the blades of at least some falcigers has been described for the type species of the genus, *S. hystrix* Claparède 1863, and for *S. boeroi* Musco, Çinar and Giangrande, 2005. Re-examination of material of *S. hystrix* revealed that some individuals –contrary to descriptions available in the literature– possess a subdistal spine not only on the blades of the falcigers in anterior but also in posterior chaetigers. Consequently, this characteristic could not be used to unambiguously distinguish the Israeli material from *S. hystrix*. In order to clarify the relationship between the three very similar species possessing falcigers with a subdistal spine, a morphometric analysis has been performed, a method allowing not only to discriminate statistically significant groupings but also to identify taxonomically important characters (Costa-Paiva and Paiva 2007).

Material and methods**Specimen collection and processing**

Specimens were collected on 11 Oct. 2009 in Haifa Bay, (Israel, Eastern Mediterranean Sea) from fine to medium sands in shallow waters (10 m). Sediment samples were taken with a Van-Veen grab (KAHLSICO, model WA265/SS214) 32×35 cm, volume 20 l, penetration 20 cm. The sediment was preserved in buffered formalin 10% for 3–7 days, then sieved through a 250 µm mesh sieve and subsequently stored in 70% ethanol. Specimens were examined under an Olympus SZx12 stereomicroscope and an Olympus BX50 microscope. Illustrations in pencil were made by means of a drawing tube, subsequently scanned, imported into a graphic program (GIMP), re-drawn

and saved as a vector graphic. Three specimens selected for obtaining Scanning Electron Microscope (SEM) images were dehydrated, critical point dried (Bal-Tec CPD 030), sputter-coated with gold (Bal-Tec SCD 050) and examined under a JEOL JSM-6390LV at the Department of Biology, University of Crete. Specimens are deposited in the invertebrate collection of the Smithsonian National Museum of Natural History, Washington D.C., USA (USNM) and in the Tel Aviv University Zoological Museum, Israel (TAU).

Morphometric analyses

A total of 30 individuals belonging to three species (*S. boeroi*: 3 individuals; *S. hystrix*: 21 individuals; *S. levantina* sp. n.: 6 individuals) were analysed. Twenty-five variables were measured: I. body length, to account for size-dependencies of other characters; II. number of chaetigers; III. length of blade of dorsalmost falciger of a) anterior, b) midbody, c) posterior chaetigers; IV. length of blade of ventralmost falciger of a) anterior, b) midbody, c) posterior chaetigers; V. ratio of length of blades of dorsalmost to ventralmost falciger in a) anterior, b) midbody, c) posterior chaetigers; VI. ratio of length of blades of falcigers in anterior to posterior chaetigers for a) dorsalmost; b) ventralmost falciger; VII. Ratio of length of dorsalmost falciger to body length in in a) anterior, b) midbody, c) posterior chaetigers; VIII. Ratio of length of ventralmost falciger to body length in in a) anterior, b) midbody, c) posterior chaetigers; IX. maximum length of serration of falcigerous blades in a) anterior, b) midbody, c) posterior chaetigers (smooth, finely serrated, strongly serrated); X. presence of a subdistal spine in dorsalmost falcigerous blades of in a) anterior, b) midbody, c) posterior chaetigers.

Body length was measured excluding antennae, anal cirri and palps. Falciger blade lengths were measured from point of insertion into shaft to distal tip. Falciger blade lengths could not always be measured on the same chaetiger in all animals if blades were broken. Instead, measurements were made in predefined body regions (anterior: first 1–5 chaetigers; posterior: last 5–7 chaetigers; midbody: in between). Three individuals of *S. hystrix* were excluded from the multivariate statistical analysis due to missing values for some characters.

Summary statistics (mean, minimum, maximum, standard deviation, coefficient of variation and range of values) were calculated for each species (measurements and calculations available in online supplementary material:

<http://polychaetes.marbigen.org/content/measured-values-sphaerosyllis-specimens>

<http://polychaetes.marbigen.org/content/summary-statistics-sphaerosyllis-hystrix>

<http://polychaetes.marbigen.org/content/summary-statistics-sphaerosyllis-boeroi>

<http://polychaetes.marbigen.org/content/summary-statistics-sphaerosyllis-levantina>

To take the different data types (numerical, categorical, binary) into account, Gower's similarity coefficient (Gower 1971) was chosen to calculate a similarity matrix.

Multidimensional Scaling (MDS) was subsequently employed to display the similarities of the different individuals. To test for significance of differences between species a PERMANOVA (Permutational Multivariate Analysis of Variance) was performed (Anderson 2001). A Principal Component Analysis (PCA) was used to determine variability of characters and to identify characters for the species differentiation. To determine the importance of the characters discriminating the species, the Principal Component Scores were correlated (Spearman's correlation coefficient) with the measured character values of each individual.

Multivariate statistical analyses were performed with PRIMER V6, correlation of the Principal Component Scores were calculated with the R package (R package version 2.10; <http://www.R-project.org>).

Electronic publication

The description of the new taxon was prepared in a Virtual Research Environment (Scratchpads) allowing for rapid and simultaneous publication of the results in print as well as electronically in a semantically enhanced form (Blagoderov et al. 2010, Penev et al. 2010). This publication and all supplementary data (measurements, results of statistical analyses, images) can be accessed on the Polychaete Scratchpads (<http://polychaetes.marbigen.org>).

Results

Taxonomic results

Sphaerosyllis levantina sp. n.

urn:lsid:zoobank.org:act:9CEE8F90-9596-49F6-AA22-BB79C0E816D9

http://species-id.net/wiki/Sphaerosyllis_levantina

Figures 1–4

Type material. Holotype (USNM 1160540) ALA-IL-7, Haifa Bay, 10.5 m depth. Label: “*Sphaerosyllis levantina*, Haifa Bay, coll. B. Galil 11.10.09 [Holotype]”. Paratypes USNM 1160541–1160573: 33 individuals, TAU-AN 25006: 10 individuals; Haifa Bay, Israel, Eastern Mediterranean Sea, Station ALA-IL-7, coll. 11.10.2009, depth 10.5 m; Labels: “*Sphaerosyllis levantina*, Haifa Bay, coll. B. Galil 11.10.09 [Paratype X]” (where X=1–43). All material preserved in 96% Ethanol.

Comparative material examined. *S. boeroi* Musco, Çinar, and Giangrande, 2005 (Southern Evoikos Gulf, Aegean Sea, Greece: 3 specimens [Label: Tribe *Sphaerosyllis*]). *S. hystrix* (Southern Evoikos Gulf, Aegean Sea, Greece: 1 specimen [Label: Tribe *Sphaerosyllis*]; Northern Evoikos Gulf, Aegean Sea, Greece: 7 specimens [Label: DI9a 7.3.91 *Sphaerosyllis hystrix*, checked S.Martín], all deposited the in Hel-

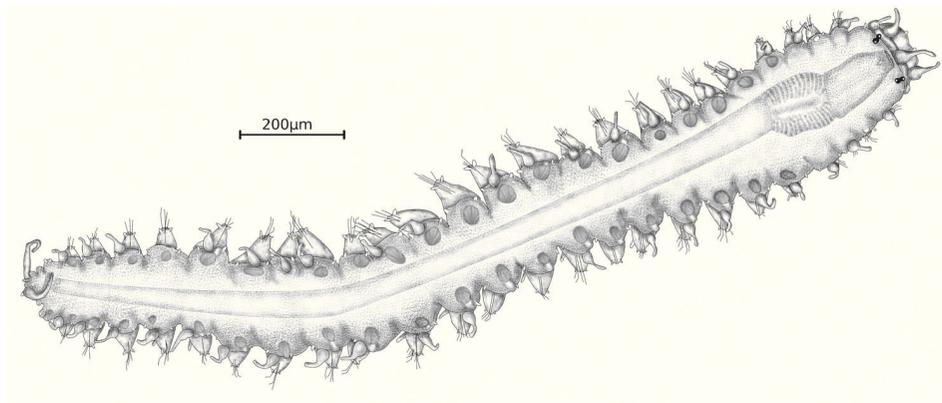


Figure 1. *Sphaerosyllis levantina* sp. n. holotype, dorsal view

lenic Centre for Marine Research, Anavyssos, Greece; Chalkida, Aegean Sea, Greece: 1 specimen [Label: 56 – *Sphaerosyllis hystrix*, κατώτερη μεσοπαραλιακή Χαλκίδας, Στενά Ευρίπου, Ξενοδοχείο Λούσι, St. 18, 25.9.97 0-0.5m, Άτομα: 1, Διδακτορικού Μίλτου] (= lower intertidal zone, Chalkida, Eviros Straight, Hotel Lousi, coll. M.S. Kitsos), Chalkida, Aegean Sea, Greece: 1 specimen [Label: 26 – *Sphaerosyllis hystrix*, κατώτερη μεσοπαραλιακή Χαλκίδας, Στενά Ευρίπου, Ξενοδοχείο Παλιρροία, St. 1a, 24.9.97 0-0.5m, Άτομα: 1, Διδακτορικού Μίλτου] (= lower intertidal zone, Chalkida, Eviros Straight, Hotel Palirroia, coll. M.S. Kitsos), Chalkida, Aegean Sea, Greece: 6 specimens [Label: 33 – *Sphaerosyllis hystrix*, κατώτερη μεσοπαραλιακή Χαλκίδας, Στενά Ευρίπου, Ξενοδοχείο Παλιρροία, St. 1α, 24.9.97 0-0.5m, Άτομα: 6, Διδακτορικού Μίλτου] (= lower intertidal zone, Chalkida, Eviros Straight, Hotel Palirroia, coll. M.S. Kitsos), Chalkida, Aegean Sea, Greece: 4 specimens [Label: 80 – *Sphaerosyllis hystrix*, κατώτερη μεσοπαραλιακή Χαλκίδας, Στενά Ευρίπου, Ξενοδοχείο Παλιρροία, St. 1α, 24.9.97 0-0.5m, Άτομα: 6, Διδακτορικού Μίλτου] (= lower intertidal zone, Chalkida, Eviros Straight, Hotel Palirroia, coll. M.S. Kitsos), Thessaloniki, Aegean Sea, Greece, 1 specimen [Label: 66 – *Sphaerosyllis hystrix*, κατώτερη μεσοπαραλιακή Λιμάνι Θεσσαλονίκης, 2γ, 6.10.97 0-0.5m, Άτομα: 1, Διδακτορικού Μίλτου] (= lower intertidal zone, Port of Thessaloniki, coll. M.S. Kitsos), all deposited in the Zoological Museum of the Aristotle University of Thessaloniki, Greece.

Type locality. Eastern Mediterranean Sea, Levantine Basin, Israel, Haifa Bay (32°54.533N, 35°04.071E).

Description. Holotype, entire animal, with 25 chaetigers, length 1.9 mm with palps but without anal cirri; width at sixth chaetiger 250 μm without parapodia, 300 μm with parapodia. Body small, slender, widest at level of proventricle (Fig. 1). Dorsal papillation on anterior chaetigers irregular, after proventricle in four longitudinal rows: two mid-dorsal rows with two papillae per segment, lateral rows with three papillae near dorsal cirri (Fig. 2a). Ventrums without visible papillation. Prostomium wider than long with 4 coalescent lensed eyes in trapezoidal arrangement. Anterior eyespots absent. Antennae pyriform with bulbous bases and elongated tips, median antenna

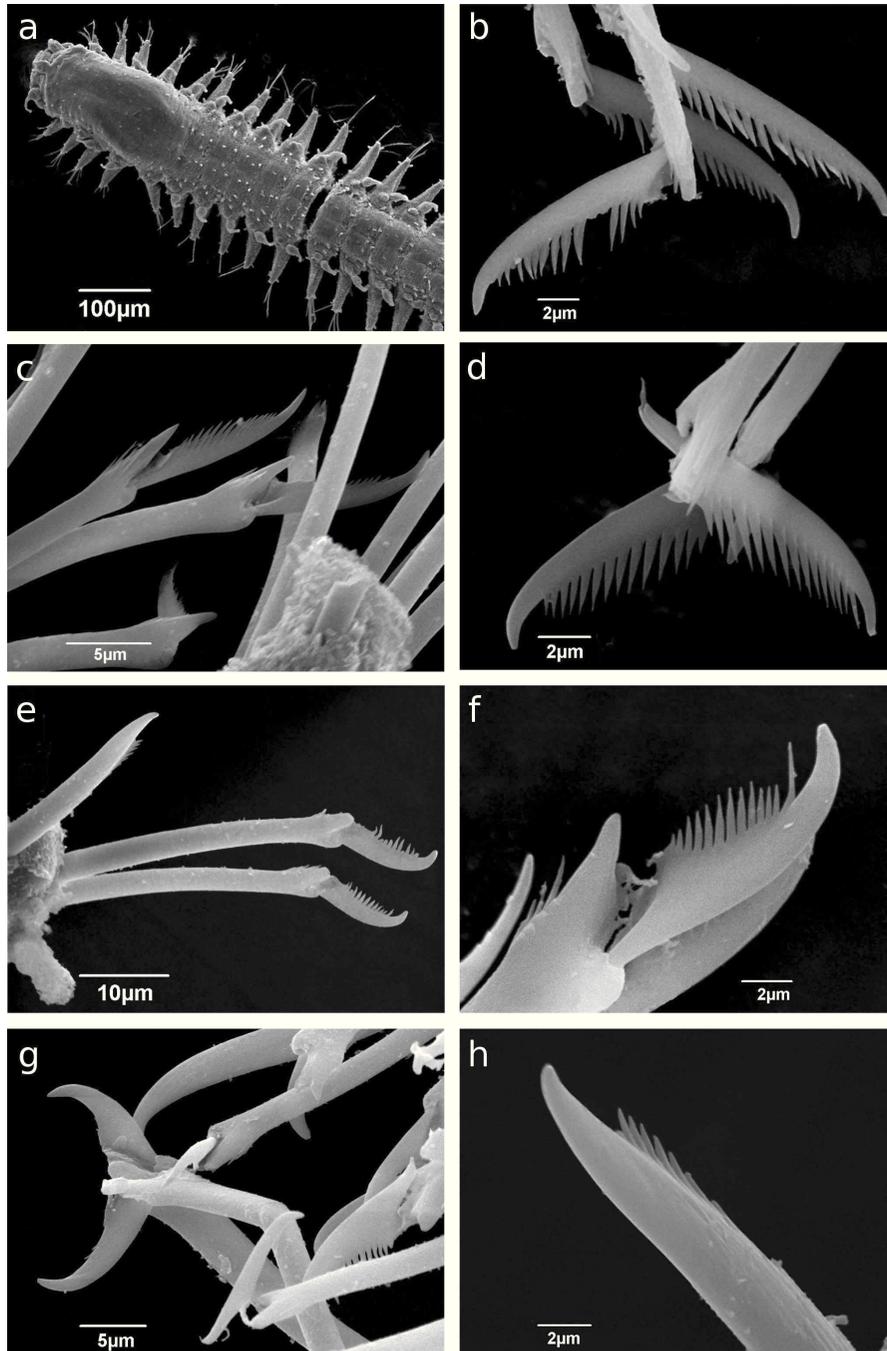


Figure 2. *Sphaerosyllis levantina* sp. n. SEM images of **a** anterior end and midbody, dorsal view **b–c** compound chaetae, anterior chaetigers **d** dorsalmost compound chaetae, anterior chaetiger **e** compound and dorsal simple chaetae, midbody **f** dorsalmost compound chaeta, posterior chaetiger **g** ventralmost compound chaetae, posterior chaetiger **h** dorsal simple chaeta

40 μm long, lateral ones 33 μm , longer than prostomium and palps together. Median antenna inserted between anterior pair of eyes, lateral ones attached on anterior margin of prostomium (Fig. 1). Palps directed ventrally, fused along their length, with a dorsal notch and few small papillae. Peristomium indistinct, dorsal fold partly covering prostomium. One pair of tentacular cirri, shaped like antennae but shorter (23 μm). Second chaetiger without dorsal cirri but with large papilla instead. Dorsal cirri similar in shape and length to tentacular cirri, anteriorly as long as parapodial lobes (23 μm), posteriorly slightly longer (28 μm). Ventral cirri conical, half as long as parapodial lobe, originating at bases of parapodia. Parapodial lobes triangular, with small papilla on each side of distal end. Parapodial glands with fibrillar material and with conical opening; from fourth chaetiger. Anterior parapodia with 4–5, rarely with 6 falcigers per fascicle; blades slender, unidentate with small subdistal spine and strong serration on 1–2 dorsalmost falcigers (Figs 2b–d, 3a). Dorso-ventral gradation in length of blades, dorsal ones maximally 14 μm , ventral ones 10 μm . Posteriorly, dorsal blades of similar length (13 μm), but stouter and more curved with robust subdistal spine and strong serration as long as subdistal spine (Figs 2e, f, 3b, c). Dorsalmost falciger posteriorly thicker than remaining ones in fascicle. Blades of ventral falcigers similar throughout body (Fig. 2g). All shafts with fine serration (Fig. 2c). Dorsal simple chaeta from chaetiger 1, subdistally serrated (Figs 2h, 4a). Ventral simple chaeta on posterior chaetigers, sigmoid, smooth (Fig. 4b). Anteriorly two aciculae per parapodium, one distally bent at right angle, acuminate tip curved upwards, the other straight and blunt (Fig. 4c); posteriorly only one acicula of the former type per parapodium. Pharynx occupying three chaetigers. Width more than $\frac{3}{4}$ of width of proventricle. Pharyngeal tooth located on anterior margin, surrounded by a crown of soft papillae. Proventricle in chaetigers 3–4 (120 μm long) with 15–17 muscle cell rows. Pygidium papillated, with two cirriform anal cirri twice as long as dorsal cirri (60 μm) (Fig. 1).

Etymology. Derived from the type locality (Levantine Basin), *levantina* being a neo-Latin adjective meaning “pertaining to the region where the sun raises”; feminine declination in accordance with the genus name (*Syllis* was a river nymph in the greek mythology and thus female).

Distribution. Israeli Coast (Levantine Basin, Eastern Mediterranean Sea).

Habitat. Fine to medium sands.

Taxonomic remarks. *S. levantina* sp. n. is similar to *S. minima* Hartmann-Schröder, 1960 in having blades of falcigers with strong serration throughout the body. However, *S. minima* has a stronger dorso-ventral gradation of the blades of falcigers (dorsal ones twice as long as ventral ones) than *S. levantina* sp. n. (dorsal ones 1.5 times longer than ventral ones) and the ventral cirrus is longer than the parapodial lobe in *S. minima*, whereas is half as long as the parapodial lobe in *S. levantina* sp. n. *S. capensis* Day, 1953, *S. taylori* Perkins, 1981, and *S. sandrae* Álvarez and San Martín, 2009 are similar to *S. levantina* sp. n. in the shape and serration of the blades of the falcigers, but *S. capensis* has all antennae positioned in line (median one posteriorly of lateral ones in *S. levantina* sp. n.), *S. taylori* shows no dorso-ventral gradation of the falciger blade length (dorsal blade 1.5 times longer than ventral one in *S. levantina* sp. n.) and

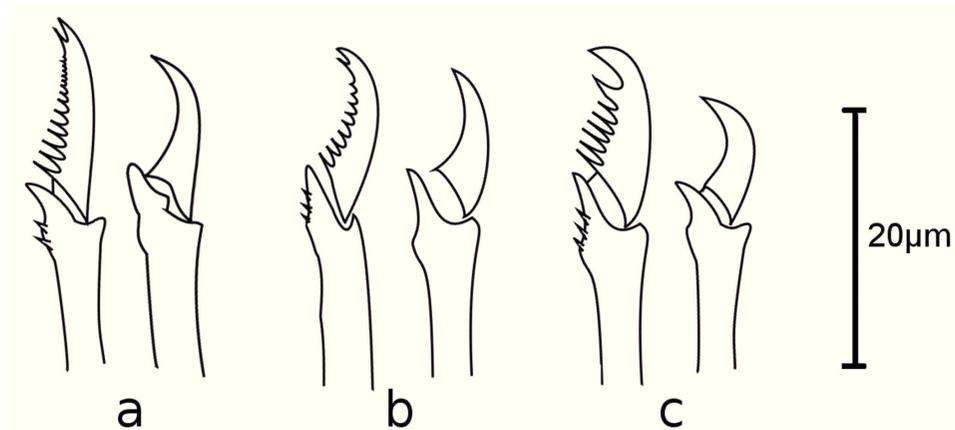


Figure 3. *Sphaerosyllis levantina* sp. n. Dorsal (left) and ventral (right) falciger of **a** anterior **b** midbody **c** posterior chaetiger

S. sandrae has smooth falcigerous blades posteriorly and parapodial glands with hyaline material (strongly serrated blades throughout the body and parapodial glands with fibrillar material in *S. levantina* sp. n.). All the above species differ from *S. levantina* sp. n. by lacking a subdistal spine on the blades of the falcigers. The only *Sphaerosyllis* species known to possess this spine are *S. hystrix* Claparède, 1863, *S. parabulbosa* San Martín and López, 2002 and *S. boeroi* Musco Çinar and Giangrande, 2005. *S. parabulbosa* clearly differs from *S. levantina* sp. n. by having minute dorsal cirri and antennae, by the presence of a subdistal spine only on blades of the posterior falcigers and by smooth blades of posterior falcigers. *S. boeroi* differs from *S. levantina* sp. n. in having much longer blades of the falcigers which show a more pronounced dorso-ventral gradation (dorsal blades 2.6 times longer than ventral ones in *S. boeroi*, 1.5 times longer in *S. levantina* sp. n.) than those of *S. levantina* sp. n. (Figs 3, 5, see also tables in online supplementary material), by having a subdistal spine on blades of all falcigers (only on dorsalmost ones in *S. levantina* sp. n.) and by the dorsalmost falcigers being serrated only proximally. *S. hystrix*, according to the literature, has a subdistal spine only on the blades of the anterior dorsalmost falcigers. However, in the examined material of *S. hystrix* from the Aegean Sea 8 out of 21 specimens also possessed a subdistal spine in posterior falcigers. *S. hystrix* can nevertheless be distinguished from *S. levantina* sp. n. by having smooth or finely serrated posterior falcigers (serration less than half the length of the subdistal spine), even when the spine is present (serration almost as long as subdistal spine in *S. levantina* sp. n.) (Figs 2f, 3, 6). Furthermore, the blades of the dorsalmost falcigers show an anteroposterior gradation in length in *S. hystrix* (anteriorly 1.5 times longer than posteriorly), whereas they are of similar length throughout the body in *S. levantina* sp. n. (Figs 3, 6, see also tables in online supplementary material). Finally, *S. hystrix* has a very narrow pharynx (almost half the width of proventricle), whereas the pharynx of *S. levantina* sp. n. is wider than $\frac{3}{4}$ of the width

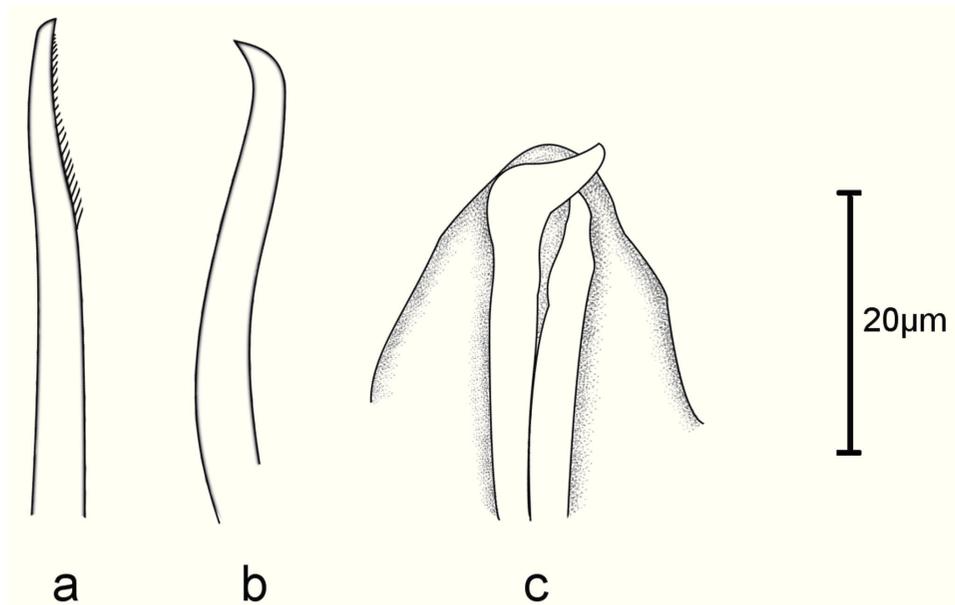


Figure 4. *Sphaerosyllis levantina* sp. n. **a** dorsal **b** ventral simple chaeta **c** aciculae, anterior chaetiger

of the proventricle. An identification key to the Mediterranean *Sphaerosyllis* species is provided at the end of this manuscript.

Ben-Eliahu (1977) discusses two different morphological forms of *S. hystrix* occurring in her samples from Israel. Based on her description and illustrations, the animal identified as *S. hystrix* sensu Westheide 1974 could potentially belong to *S. levantina* sp. n. because of the similar characters of falcigers and papillation. However, the description does not report the characteristic subdistal spine on the blades of the posterior falcigers. In addition, Westheide's (1974) description of *S. hystrix* from the Galápagos Islands differs from both Ben-Eliahu's specimen and the present material by the absence of parapodial glands (Westheide 1974), a character considered as variable and thus of no taxonomic value by Ben-Eliahu (1977) but recently accepted as a taxonomically stable character (Riser 1991).

Multivariate morphometrical analysis

The results of the Principal Component Analysis show that the first principal component (PC1) account for 77.4% of the variability, the second (PC2) for 16.4% and the remaining 3 PCs for 5.1% (eigenvector values available at <http://polychaetes.marbigen.org/content/morphometric-analysis-pca-eigenvectors>). The Spearman's correlation of the Principal Component scores with the measured character values of the individuals revealed that the length of the dorsalmost falcigerous blades in all

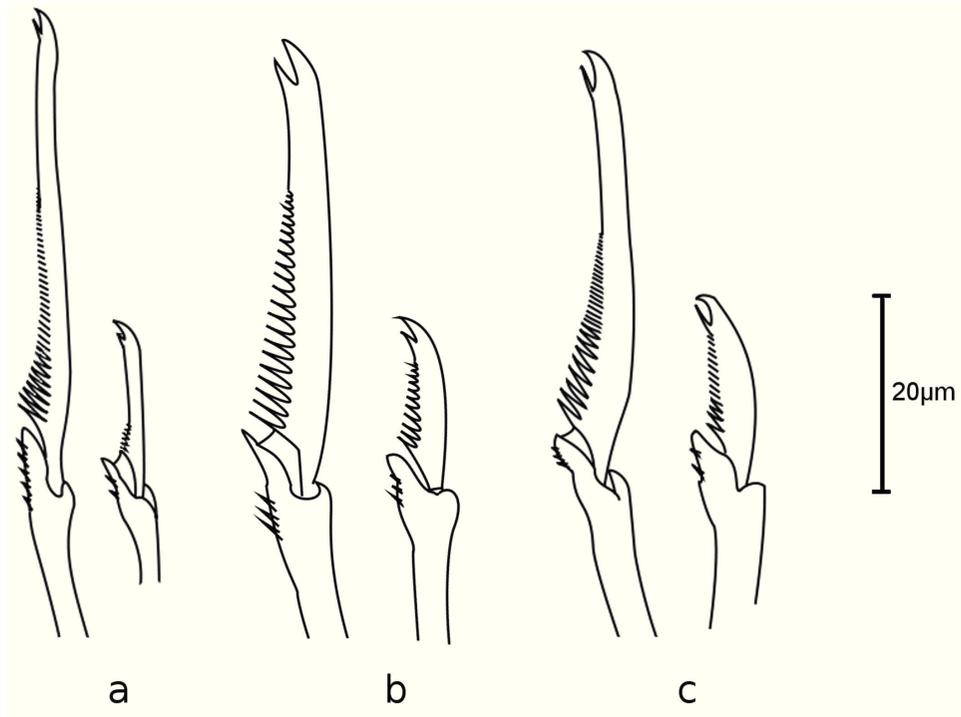


Figure 5. *Sphaerosyllis boeroi*. Dorsal (left) and ventral (right) falciger of **a** anterior **b** midbody **c** posterior chaetiger

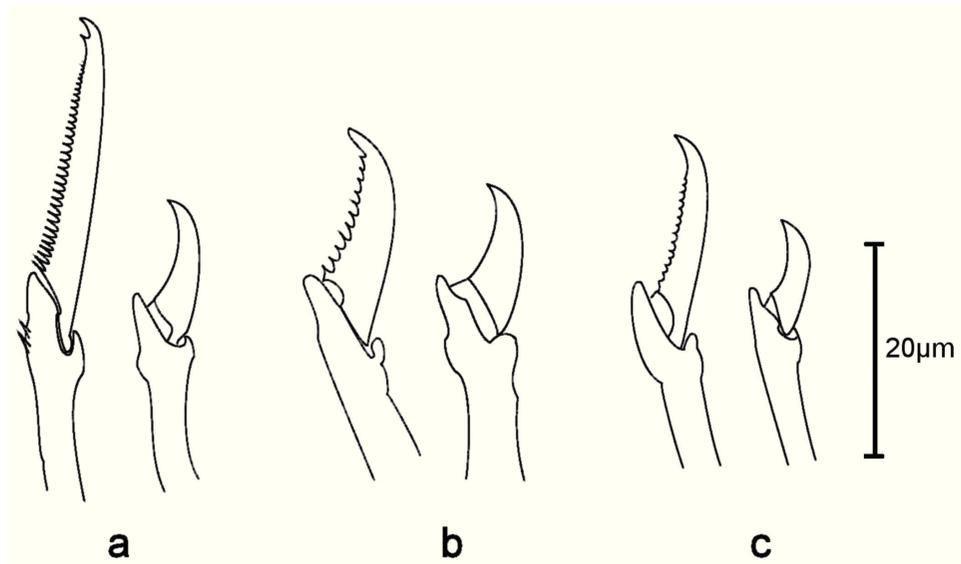


Figure 6. *Sphaerosyllis hystrix*. Dorsal (left) and ventral (right) falciger of **a** anterior **b** midbody **c** posterior chaetiger

body parts (anterior, midbody, posterior), as well as the ratio of the anterior to posterior ventralmost falcigerous blade are the most important characters discriminating between the three species (Q -values >0.8 / <-0.8 at $p < 0.005$) (<http://polychaetes.marbigen.org/content/spearman-correlation-principal-component-scores-vs-measurements>).

The PCA plot of the first two components show a discrimination of species into three groups, with individuals of *S. levantina* sp. n. having the lowest PC1 scores, *S. boeroi* the highest scores. *S. levantina* sp. n. and *S. hystrix* show similar PC2 scores, whereas *S. boeroi* shows lower scores, and, except for one small-sized individual, forms a distinct group apart from the remaining species. Individuals of *S. levantina* sp. n. likewise form a close group, however, a couple of individuals of *S. hystrix* cannot be distinguished from this cluster (Fig. 7). The MDS diagram gives similar results, with individuals of *S. boeroi* and *S. levantina* sp. n. forming distinct groups, whereas individuals of *S. hystrix* are spread as a heterogeneous group, with some of them being plotted close to individuals of either *S. boeroi* or *S. levantina* sp. n. (Fig. 8).

The PERMANOVA analysis results in a p -value of 0.001 as calculated by 999 permutations, thus the null-hypothesis (no differences between the groups) cannot be sustained. Subsequent analyses of the differences between species through pairwise tests reveals significant differences between species (*S. hystrix* / *S. boeroi*: $p = 0.003$, 713 permutations; *S. hystrix* / *S. levantina* sp. n.: $p = 0.001$, 995 permutations; *S. boeroi* / *S. levantina* sp. n.: $p = 0.015$, 84 permutations).

Discussion

The genus *Sphaerosyllis*—like many of the small-sized Exogoninae genera—has a difficult and often confused taxonomy and biogeography. Among the potential causes contributing to the current confusion the following could be cited: a) lack of detail in older (before ca. 1970) species descriptions; b) difficulties of observing certain characters in fixed material (Riser 1991); c) descriptions of new species without examination of comparative material; d) ongoing discussions on the taxonomic value of characters such as the presence or absence of dorsal cirri on the second chaetiger (Fauvel 1923, San Martín 2005), presence and type of parapodial glands (Westheide 1974, Ben-Eliahu 1977, Riser 1991) and variations in chaetal structures (Riser 1991). These factors have led to the assignment of individuals with very different character sets to the same species name and thus to wide-spread distribution records of some species. *S. hystrix* (type locality Normandy, France) is included among those species with an alleged cosmopolitan distribution, since it has been recorded from most European coasts including the Mediterranean Sea, the north-western coasts of America (Berkeley and Berkeley 1948, Hartman 1968), the Galápagos Islands (Westheide 1974), China (Men et al. 1993, Ding and Westheide 2008), Australia (Hartmann-Schröder 1984, 1985) and the Western Atlantic (Hartman and Fauchald 1971, Temperini 1981), among others. However, recent studies suggest that the North American records of *S. hystrix* and *S. pirifera* Claparède, 1868 are in fact

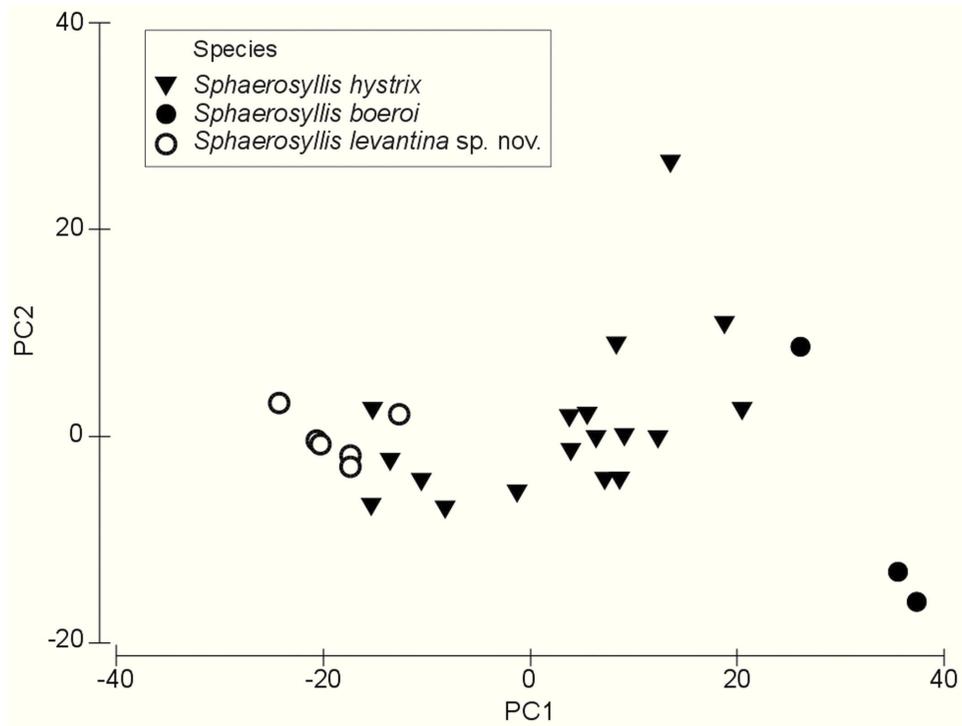


Figure 7. PCA plot.

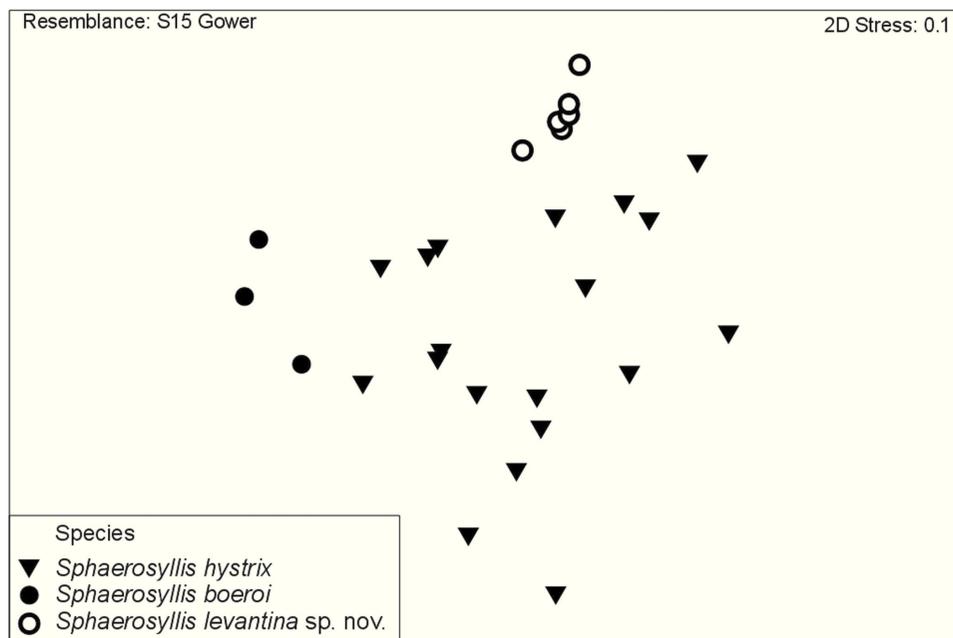


Figure 8. MDS plot.

individuals of *S. californiensis* Hartman, 1966 and that the two European species are not represented in the American Pacific fauna (Kudenov and Harris 1995). Similarly, some specimens from the Mediterranean Sea previously identified as *S. hystrix* had been re-examined and found to exhibit significant morphological differences to *S. hystrix*, leading to the establishment of a new species, *S. boeroi* (Musco et al. 2005). In the light of an ever-increasing number of molecular analyses revealing cryptic species complexes in morphologically indistinguishable polychaete species with an assumed cosmopolitan distribution (e.g. Westheide and Hass-Cordes 2001, Westheide and Schmidt 2003, Barroso et al. 2009, Bleidorn et al. 2006) it is likely that the various specimens recorded under the name *S. hystrix* may in fact form a complex of similar species, especially since many descriptions differ substantially from each other (see Ben-Eliahu 1977).

The morphometric analysis conducted in this study support the hypothesis of several morphologically very similar species co-existing in the Mediterranean. The individuals of *S. levantina* sp. n. and *S. boeroi* form distinct groups in the PCA and MDS plots, however the individuals of *S. hystrix* show a much wider spread, marginally overlapping with the other two species when only the meristic characters are taken into account. This is explained through a high character variability in the examined individuals, especially concerning the presence of a subdistal spine on the blades of the posterior falcigers and the length of the falciger blades. The presence of a subdistal spine on all dorsal falcigerous blades is invariable in *S. boeroi* and *S. levantina* sp. n., whereas individuals of *S. hystrix* with otherwise very similar chaetal structures might or might not possess such spine. Another feature that seems to be highly variable in *S. hystrix* is the length of the falciger blades in relation to body size. In fact, individuals of *S. levantina* sp. n. with short falciger blades are located at the lower end of the size spectrum of all measured blades, *S. boeroi* with almost spiniger-like blades at the higher end, whereas the blade lengths of the examined individuals of *S. hystrix* form a smooth transition between the other two species.

However, when tested by strict statistical criteria, the hypothesis of different co-existing species is significantly supported, and based on their meristic characters the species show significant differences. The results of the current study suggest that *S. hystrix* may well constitute a species complex. Given the difficult taxonomic status of the genus, similar results might be expected for other species as well, and consequently, distributions of several *Sphaerosyllis* species might be in fact questionable or unknown.

Key to the Mediterranean *Sphaerosyllis* species:

The three species *S. claparedei* Ehlers, 1864, *S. papillifera* Naville, 1933 and *S. ovigera* Langerhans, 1879 are poorly known. All have been described as having dorsal cirri on the second chaetiger, however, other species, such as *S. hystrix*, were also originally described or illustrated with dorsal cirri on the second chaetiger whereas they are in fact absent. Since the three aforementioned species are exclusively known from their

original description (or partly reproductions of these) and have never been re-described based on new material, they are tentatively included in the key below, but their identity remains questionable.

- | | | |
|----|---|---|
| 1 | Dorsal cirri on chaetiger 2 present | 2 |
| – | Dorsal cirri on chaetiger 2 absent..... | 4 |
| 2 | Papillae on dorsum absent | <i>Sphaerosyllis claparedei</i> Ehlers, 1864 |
| – | Papillae on dorsum present | 3 |
| 3 | Parapodial glands absent | <i>Sphaerosyllis papillifera</i> Naville, 1933 |
| – | Parapodial glands with fibrillar material | <i>Sphaerosyllis ovigera</i> Langerhans, 1879 |
| 4 | Parapodial glands present..... | 5 |
| – | Parapodial glands absent | 15 |
| 5 | Parapodial glands with fibrillar material | 6 |
| – | Parapodial glands with granular material..... | 12 |
| 6 | All antennae in line..... | <i>Sphaerosyllis capensis</i> Day, 1953 |
| – | Median antenna inserted more posteriorly than lateral ones..... | 7 |
| 7 | Dorsal cirri shorter than parapodial lobes, at least in anterior chaetigers..... | 8 |
| – | Dorsal cirri longer than parapodial lobes..... | 9 |
| 8 | Blades of falcigers strongly serrated, short (<10µm); shafts with strong spines..... | <i>Sphaerosyllis thomasi</i> San Martín, 1984 |
| – | Blades of falcigers with serration only anteriorly and dorsalmost; blades with slight dorso-ventral gradation but always longer than 10µm; shafts smooth... .. | <i>Sphaerosyllis parabolbosa</i> San Martín and López, 2002 |
| 9 | Blades of falcigers without marked dorso-ventral gradation in length..... | <i>Sphaerosyllis taylori</i> San Martín, 1984 |
| – | Blades of dorsalmost falcigers at least 1.5 times the length of ventral ones.... | 10 |
| 10 | Blades of posterior dorsal compound falcigers smooth to finely serrated | <i>Sphaerosyllis hystrix</i> Claparède, 1863 |
| – | Blades of posterior dorsal compound falcigers strongly serrated (spinules of almost same length as the subdistal spine)..... | 11 |
| 11 | Blades of anterior dorsal compound falcigers at least twice as long as ventral ones; anteroposterior gradation of blade length; blades of both dorsal and ventral compound chaetae with a subdistal spine | <i>Sphaerosyllis boeroi</i> Musco, Çinar and Giangrande, 2005 |
| – | Blades of anterior dorsal compound falcigers less than twice as long as ventral ones; no anteroposterior gradation of blade length; blades of only dorsal compound chaetae with a subdistal spine | <i>Sphaerosyllis levantina</i> sp. n. |
| 12 | Blades of dorsalmost falcigers long (>30µm), at least twice as long as ventral ones | <i>Sphaerosyllis magnidentata</i> Perkins, 1981 |
| – | Blades of falcigers short (<15µm), with only slight dorso-ventral gradation... .. | 13 |
| 13 | Dorsal cirri clearly longer than parapodial lobes..... | <i>Sphaerosyllis</i> sp. [San Martín 2003] |

	Sphaerosyllis levantina sp. n. (Annelida) from the eastern Mediterranean...	341
–	Dorsal cirri as long as or shorter than parapodial lobe.....	14
14	Antennae bulbous with small tip, shorter than prostomium; dorsal simple chaetae smooth; anterior parapodia with two aciculae, one straight, one with tip bent at right angle	
 <i>Sphaerosyllis</i> sp. Del-Pilar-Ruso & San Martín, in press	
–	Antennae pyriform, as long as prostomium, dorsal simple chaetae serrated, all parapodia with one acicula.....	<i>Sphaerosyllis glandulata</i> Perkins, 1981
15	All aciculae straight.....	16
–	Tip of some aciculae bent at right angle.....	17
16	Dorsal cirri with conspicuous papilla, giving cirri a bifid appearance.....	
 <i>Sphaerosyllis gravinae</i> Somaschini & San Martín, 1994	
–	Dorsal cirri without papilla	<i>Sphaerosyllis bulbosa</i> Southern, 1914
17	All antennae in line.....	<i>Sphaerosyllis austriaca</i> Banse, 1959
–	Median antenna inserted more posteriorly than lateral ones.....	18
18	Anterior parapodia with two aciculae, one straight, one with tip bent at right angle; pharyngeal glands on chaetiger 1 present	
 <i>Sphaerosyllis pirifera</i> Claparède, 1868	
–	All parapodia with one acicula only; pharyngeal glands on chaetiger 1 absent	<i>Sphaerosyllis piriferopsis</i> Perkins, 1981

Acknowledgements

The authors kindly acknowledge assistance from the following colleagues: Dr Nomiki Simboura (HCMR, Anavyssos) and Dr Miltiadis Kitsos (Aristotelian University of Thessaloniki) for loans of comparative material of *Sphaerosyllis hystrix* and *Sphaerosyllis boeroi*, Dr Guillermo San Martín (Universidad Autónoma de Madrid) for the provision of some of the literature resources, Mrs Alexandra Siakouli (University of Crete) for taking SEM pictures, Dr Vincent Smith, Simon Rycroft, Ben Scott (NHM, London) and Dr Lyubomir Penev (Pensoft Publishers, Bulgaria) for support with the electronic Scratchpads publication. The two reviewers (Dr Guillermo San Martín, Dr Luigi Musco) are thanked for their suggestions to improve the manuscript. Financial support was provided by the ViBRANT project (FP7, EU).

References

- Anderson M (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26: 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x
- Álvarez P, San Martín G (2009) A new species of *Sphaerosyllis* (Annelida: Polychaeta: Syllidae) from Cuba, with a list of syllids from the Guanahacabibes Biosphere Reserve (Cuba). *Journal of the Marine Biological Association of the UK* 89: 1427–1435. doi: 10.1017/S0025315409000654

- Banse K (1959) Über die Polychaeten-Besiedlung einiger submariner Höhlen. Ergebnisse der Österreichischen Tyrrheni- Expedition 1952, Teil XII. Pubblicazioni della Stazione Zoologica di Napoli 30: 417–469.
- Barroso R, Klautau M, Solé-Cava AM, Paiva PC (2009) *Eurythoe complanata* (Polychaeta: Amphinomidae), the “cosmopolitan” fireworm, consists of at least three cryptic species. Marine Biology 157: 69–80. doi: 10.1007/s00227-009-1296-9
- Ben-Eliahu MN (1977) Polychaete cryptofauna from rims of similar intertidal vermetid reefs on the Mediterranean coast of Israel and in the Gulf of Elat: Exogoninae and Autolytinae (Polychaeta Errantia: Syllidae). Israel Journal of Zoology 26: 59–99.
- Berkeley E, Berkeley C (1948) Canadian Pacific Fauna. Fisheries Research Board of Canada, Toronto, 100 pp.
- Blagoderov V, Brake I, Georgiev T, Penev L, Roberts D, Rycroft S, Scott B, Agosti D, Catapano T, Smith VS (2010) Streamlining taxonomic publication: a working example with Scratchpads and ZooKeys. ZooKeys 50: 17–28. doi: 10.3897/zookeys.50.539
- Bleidorn C, Kruse I, Albrecht S, Bartolomaeus T (2006) Mitochondrial sequence data expose the putative cosmopolitan polychaete *Scoloplos armiger* (Annelida, Orbiniidae) as a species complex. BMC Evolutionary Biology 6: 47. doi: 10.1186/1471-2148-6-47
- Claparède É (1863) Beobachtungen über Anatomie und Entwicklungsgeschichte wirbelloser Thiere an der Küste von Normandie angestellt. Wilhelm Engelmann Verlag, Leipzig, 172pp. <http://biodiversitylibrary.org/item/40310>
- Claparède É (1868) Les annélides chétopodes du Golfe de Naples. Mémoires de la Société de physique et d’histoire naturelle de Genève 20: 313–584. <http://biodiversitylibrary.org/item/18576>
- Costa-Paiva EM, Paiva PC (2007) A morphometric analysis of *Eunice* Cuvier (Annelida, Polychaeta) species. Revista Brasileira de Zoologia 24: 353–358. doi: 10.1590/S0101-81752007000200013
- Day JH (1953) The Polychaete fauna of South Africa. Part 2. Errant species from Cape shores and estuaries. Annals of the Natal Museum 12: 397–441.
- Del-Pilar-Ruso Y, San Martín G (in press) Description of a new species of *Sphaerosyllis* Claparède, 1863 (Polychaeta: Syllidae: Exogoninae) from the Alicante coast (W Mediterranean) and first report for the Mediterranean Sea and the Iberian Peninsula of two other species of Syllidae. Mediterranean Marine Science.
- Ding Z-hu, Westheide W (2008) Interstitial Exogoninae from the Chinese coast (Polychaeta, Syllidae). Senckenbergiana Biologica 88: 125–159.
- Ehlers E (1864) Die Borstenwürmer (Annelida Chaetopoda) nach systematischen und anatomischen Untersuchungen dargestellt. Wilhelm Engelmann Verlag, Leipzig, 748 pp. <http://biodiversitylibrary.org/item/18348>
- Fauvel P (1923) Polychètes errantes. Faune de France, Paris, 488 pp. [http://www.fauvedefrance.org/bibliotheque/docs/P.FAUVEL\(FdeFr05\)Polychetes-errantes.pdf](http://www.fauvedefrance.org/bibliotheque/docs/P.FAUVEL(FdeFr05)Polychetes-errantes.pdf)
- Gower J (1971) A General Coefficient of Similarity and Some of Its Properties. Biometrics 27: 857–874. <http://www.jstor.org/pss/2528823>
- Hartman O (1966) Quantitative survey of the benthos of San Pedro Basin, southern California. Part II. Final results and conclusions. Allan Hancock Pacific Expeditions 19: 187–455.

- Hartman O (1968) Atlas of the errantiate polychaetous annelids from California. Allan Hancock Foundation, Los Angeles, 828 pp.
- Hartman O, Fauchald K (1971) Deep-water benthic polychaetous annelids off New England to Bermuda and other North Atlantic areas. Part II. Allan Hancock Monographs in Marine Biology 6: 1–327. <http://hdl.handle.net/10088/3458>
- Hartmann-Schröder G (1960) Polychaeten aus dem Roten Meer. Kieler Meeresforschungen 16: 69–125.
- Hartmann-Schröder G (1984) Die Polychaeten der antiborealen Südküste Australiens (zwischen Albany im Westen und Ceduna im Osten). Teil 10. Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut 81: 7–62.
- Hartmann-Schröder G (1985) Die Polychaeten der antiborealen Südküste Australiens (zwischen Port Lincoln im Westen und Port Augusta im Osten). Teil 11. Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut 82: 61–99.
- Kudenov JD, Harris LH (1995) Family Syllidae Grube, 1850. In: Blake et al. (Eds) Taxonomic Atlas of the Benthic Fauna of the Santa Maria Basin and Western Santa Barbara Channel. The Annelida Part 2 - Polychaeta: Phyllodocida (Syllidae and Scale-Bearing Families), Amphinomida and Eunicida. Santa Barbara Museum of Natural History, Santa Barbara, 1–97.
- Langerhans P (1879) Die Wurmfauna von Madeira. Zeitschrift für wissenschaftliche Zoologie 32: 513–592.
- Men F, Ding Z-hu, Zhao J, Wu B-L (1993) A preliminary study on small syllides from the Huanghai Sea (Yellow Sea). Journal of Oceanography of Huanghai and Bohai Seas 11: 19–36.
- Musco L, Giangrande A (2005) Mediterranean Syllidae (Annelida: Polychaeta) revisited: biogeography, diversity and species fidelity to environmental features. Marine Ecology Progress Series 304: 143–153. doi: 10.3354/meps304143
- Musco L, Çinar ME, Giangrande A (2005) A new species of *Sphaerosyllis* (Polychaeta, Syllidae, Exogoninae) from the coasts of Italy and Cyprus (eastern Mediterranean Sea). Italian Journal of Zoology 72: 161–166. doi: 10.1080/11250000509356666
- Naville A (1933) Quelques formes épitoques d'annelides polychètes nouvelles ou peu connues pechées à la lumière dans la baie de Banyuls. Annales des sciences naturelles 16: 171–208.
- Penev L, Agosti D, Georgiev T, Catapano T, Miller J, Blagoderov V, Roberts D, Smith VS, Brake I, Rycroft S, Scott B, Johnson NF, Morris RA, Sautter G, Chavan V, Robertson T, Remsen D, Stoev P, Parr C, Knapp S, Kress WJ, Thompson FC, Erwin T (2010) Semantic tagging of and semantic enhancements to systematics papers: ZooKeys working examples. ZooKeys 50: 1–16. doi: 10.3897/zookeys.50.538
- Perkins TH (1981) Syllidae, principally from Florida, with descriptions of a new genus and twenty-one new species. Proceedings of the Biological Society of Washington 93: 1080–1172. <http://biostor.org/reference/73932>
- Riser NW (1991) An evaluation of taxonomic characters in the genus *Sphaerosyllis* (Polychaeta: Syllidae). Ophelia supplement: 209–218.
- San Martín G (1984) Estudio biogeográfico, faunístico y sistemático de los poliquetos de la familia Silidos (Syllidae: Polychaeta) en Baleares. PhD Thesis, Madrid, Spain: Universidad Complutense de Madrid.

- San Martín G (2003) Annelida, Polychaeta II: Syllidae. In: Ramos MA et al. (Eds) Fauna Ibérica, vol. 21. Museo Nacional de Ciencias Naturales. CSIC, Madrid, 554 pp.
- San Martín G (2005) Exogoninae (Polychaeta: Syllidae) from Australia with the description of a new genus and twenty-two new species. *Records of the Australian Museum* 57: 39–152. doi: 10.3853/j.0067-1975.57.2005.1438
- San Martín G, López E (2002) New species of *Autolytus* Grube, 1850, *Paraprocerastea* San Martín & Alós, 1989, and *Sphaerosyllis* Claparède 1863 (Syllidae, Polychaeta) from the Iberian Peninsula. *Sarsia* 87: 135–143. doi: 10.1080/003648202320205210
- Somaschini A, San Martín G (1994) Description of two new species of *Sphaerosyllis* (Polychaeta: Syllidae: Exogoninae) and first report of *Sphaerosyllis glandulata* for the Mediterranean Sea. *Cahiers de Biologie Marine* 35: 357–367.
- Southern R (1914) Clare Island Survey. Archiannelida and Polychaeta. *Proceedings of the Royal Irish Academy* 31: 1–160. <http://biodiversitylibrary.org/page/34773787>
- Temperini MI (1981) “Sistemática e Distribuição dos Poliquetos Errantes da Plataforma Continental Brasileira entre as Latitudes de 23°05'S" e 30°00'S". MSc Thesis, São Paulo, Brazil: Universidade de São Paulo.
- Westheide W (1974) Interstitielle Fauna von Galapagos. XI. Pisionidae, Hesionidae, Pilargidae, Syllidae (Polychaeta). *Mikrofauna des Meeresbodens* 44: 1–146.
- Westheide W, Hass-Cordes E (2001) Molecular taxonomy: description of a cryptic *Petitia* species (Polychaeta: Syllidae) from the Island of Mahé (Seychelles, Indian Ocean) using RAPD markers and ITS2 sequences. *Journal of Zoological Systematics and Evolutionary Research* 39: 103–111. doi: 10.1046/j.1439-0469.2001.00166.x
- Westheide W, Schmidt H (2003) Cosmopolitan versus cryptic meiofaunal polychaete species: an approach to a molecular taxonomy. *Helgoland Marine Research* 57: 1–6. doi: 10.1007/s10152-002-0114-2

Supplementary material

The Scratchpads version of this publication is available at:

<http://polychaetes.marbigen.org/node/35>

Character matrices for specimens used in the morphological analysis are available at:

<http://polychaetes.marbigen.org/content/measured-values-sphaerosyllis-specimens>

Summary statistics for the species are available at:

<http://polychaetes.marbigen.org/content/summary-statistics-sphaerosyllis-hystrix>

<http://polychaetes.marbigen.org/content/summary-statistics-sphaerosyllis-boeroi>

<http://polychaetes.marbigen.org/content/summary-statistics-sphaerosyllis-levantina>

Results of the morphometric analysis are available at:

<http://polychaetes.marbigen.org/content/morphometric-analysis-pca-eigenvectors>

Sphaerosyllis levantina *sp. n.* (*Annelida*) from the eastern Mediterranean... 345

<http://polychaetes.marbigen.org/content/spearman-correlation-principal-component-scores-vs-measurements>

Illustrations and graphs of the statistical analysis are available at:

<http://polychaetes.marbigen.org/category/image-galleries/sphaerosyllis>

Chapter 3

Faulwetter S, Vasileiadou A, Kouratoras M, Dailianis T,
Arvanitidis C (2013) Micro-computed tomography:
Introducing new dimensions to taxonomy.

ZooKeys 263: 1-45.

doi: [10.3897/zookeys.263.4261](https://doi.org/10.3897/zookeys.263.4261)

Micro-computed tomography: Introducing new dimensions to taxonomy

Sarah Faulwetter^{1,2}, Aikaterini Vasileiadou^{2,3}, Michail Kouratoras⁴,
Thanos Dailianis², Christos Arvanitidis²

1 Department of Zoology-Marine Biology, Faculty of Biology, National and Kapodestrian University of Athens, Panepistimiopolis, 15784, Athens, Greece, **2** Institute for Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, 71003 Heraklion, Crete, Greece **3** Department of Biology, University of Patras, 26504, Rio, Patras, Greece **4** Hellenic Centre for Marine Research, 71003 Heraklion, Crete, Greece

Corresponding author: Sarah Faulwetter (sarifa@hcmr.gr)

Academic editor: V. Smith | Received 7 November 2012 | Accepted 28 January 2013 | Published 4 February 2013

Citation: Faulwetter S, Vasileiadou A, Kouratoras M, Dailianis T, Arvanitidis C (2013) Micro-computed tomography: Introducing new dimensions to taxonomy. *ZooKeys* 263: 1–45. doi: 10.3897/zookeys.263.4261

Abstract

Continuous improvements in the resolution of three-dimensional imaging have led to an increased application of these techniques in conventional taxonomic research in recent years. Coupled with an ever increasing research effort in cybertaxonomy, three-dimensional imaging could give a boost to the development of virtual specimen collections, allowing rapid and simultaneous access to accurate virtual representations of type material. This paper explores the potential of micro-computed tomography (X-ray micro-tomography), a non-destructive three-dimensional imaging technique based on mapping X-ray attenuation in the scanned object, for supporting research in systematics and taxonomy. The subsequent use of these data as virtual type material, so-called “cybertypes”, and the creation of virtual collections lie at the core of this potential. Sample preparation, image acquisition, data processing and presentation of results are demonstrated using polychaetes (bristle worms), a representative taxon of macro-invertebrates, as a study object. Effects of the technique on the morphological, anatomical and molecular identity of the specimens are investigated. The paper evaluates the results and discusses the potential and the limitations of the technique for creating cybertypes. It also discusses the challenges that the community might face to establish virtual collections. Potential future applications of three-dimensional information in taxonomic research are outlined, including an outlook to new ways of producing, disseminating and publishing taxonomic information.

Keywords

Micro-computed tomography, systematics, taxonomy, 3D, visualisation, interactive PDF, polychaetes, cybertypes, cybertaxonomy

Introduction

Morphology-based taxonomy has been at the heart of systematic research for over two centuries. Over the last decades, however, the dominant role of morphology in systematics and phylogenetics has been challenged by an increasing number of analyses supported by molecular data (Cook et al. 2010, Giribet 2010). Sequence data are being produced at a rapid speed and are readily available for constructing phylogenies or delimiting species. However, the formal description and naming of species and other biological units is still central to biodiversity research (Budd and Olsson 2007, Deans et al. 2011). This highly accelerated data acquisition creates an imbalance between availability of data and the human power to actually interpret them and thus to create new knowledge. In taxonomy, this “gap in scalability” (Giribet 2010) becomes even more problematic due to the time-consuming and still largely manual process of describing and naming new species which cannot keep up with the speed at which new information becomes available from the molecular world. As a consequence, a steadily increasing number of species are delimited genetically but lack a formal, morphology-based description (e.g. Audzijonyte et al. 2008, Barroso et al. 2009, Osborn and Rouse 2010). This problem is amplified by the so-called “taxonomic impediment”: fewer and fewer resources—both human and financial—are available for conventional taxonomic and systematic research (Carvalho et al. 2007), and the field is lacking a global electronic infrastructure (Godfray 2002a, Wheeler et al. 2004). The controversial debate over possible causes and remedies for the stagnation the discipline is experiencing (e.g. Knapp et al. 2002, Godfray 2002b, Carvalho et al. 2007, Evenhuis 2007, Joppa et al. 2011a, b) has stimulated governments and funding agencies to increasingly recognise its importance, and several encouraging developments have arisen over the last years. Besides releasing funds for training and education, much funding has been allocated to the field of cybertaxonomy, allowing the development of internet-based tools and resources aiming to boost taxonomic research and to accelerate the process of new species descriptions and systematic analyses. These developments include online resources such as name-based registers (e.g. Catalogue of Life – <http://www.catalogueoflife.org>, the Global Names Architecture – <http://www.globalnames.org>, ZooBank – <http://www.zoobank.org>, the World Register of Marine Species – <http://www.marinespecies.org>), biogeographic databases (e.g. the Ocean Biogeographic Information System – <http://www.iobis.org>, the Global Biodiversity Information Facility – <http://www.gbif.org>), aggregators and curators (e.g. Fishbase – <http://www.fishbase.org>, Encyclopedia of Life – <http://www.eol.org>), virtual research environments targeted towards taxonomy (Scratchpads – <http://www.scratchpads.eu>), increasing availability of literature, both for legacy literature (Biodiversity Heritage Library – <http://www.biodiversitylibrary.org>) as well as an increase in open access literature and journals aiming at rapid publication of taxonomic treatments, including new publication models such as semantically enhanced information (Penev et al. 2010).

Despite the increase in information availability through these initiatives, one of the main bottlenecks in conventional taxonomy and systematics is still the availability of type material and thus reliable morphological information (Godfray 2007, Wheeler

et al. 2012). Loans from museums are often difficult or even impossible to obtain and the time and effort to gather the material needed for a systematic revision can take months or years. The tediousness of this work contributes even further to the stagnation of the discipline: new data are produced at a slow pace, and this lack of new, readily accessible and computer-retrievable morphological data (Deans et al. 2011) prevents the testing of large-scale hypotheses, as they currently become common in phylogenomic analyses (Dunn et al. 2008, Philippe et al. 2009, Edgecombe et al. 2011). Godfray (2007) has argued that a possible solution to overcoming this bottleneck could be a web-based taxonomy using “cybertypes”, “based on the very best current imaging methods” (Godfray 2007), an idea immediately criticised by others (e.g. Carvalho et al. 2007). The idea of creating virtual collections of taxonomic material is, however, indeed enticing, and first implementations of accurate imaging methods, mass digitisations and remote access to digital material have been recently presented in a dedicated collection of papers (Smith and Blagoderov 2012, and references therein). Technological advances and a new generation of imaging techniques will inevitably open new horizons not only by providing rapid access to first-hand morphological information but also by making this information accessible to humans and computers alike. Non-destructive three-dimensional imaging techniques such as confocal laser scanning microscopy (cLSM), optical projection tomography (OPT), magnetic resonance imaging (MRI) and micro-computed tomography (micro-CT), allow for rapid creation of high-resolution morphological and anatomical data in three dimensions (Giribet 2010, Ziegler et al. 2008, 2010, 2011a, for a detailed overview of the different techniques see e.g. Boistel et al. 2011, Laforsch et al. 2012). These techniques allow detailed virtual reconstructions of the morphology and anatomy of specimens and subsequent interactive manipulation (e.g. rotation, virtual dissection) and analysis of these data. Not only are they ideal for the digitisation of taxonomically important morphological information, but they allow new kinds of analyses (e.g. morphometrics in three dimensions) to be performed, thus creating novel directions of research. Indeed, the ability of these technologies to create three-dimensional, interactive models with a resolution in the micrometre scale or even below, combined with non-destructive sample assessment (as opposed to three-dimensional reconstruction of histological sections), has lately started to attract the attention of researchers beyond the traditional (clinical) applications of the methods, boosted by an increasing accessibility of micro-CT scanners and rapid computational advances. Particularly, invertebrate zoologists have started to employ micro-MRI (for an overview of taxa imaged so far with MRI see Ziegler et al. 2011a) and micro-CT. Several studies already show the potential of these methods to deliver new data to test taxonomic hypotheses (Heim and Nickel 2010, McPeck et al. 2011, Csösz 2012). They also provide new insights into morphology and anatomy (Golding and Jones 2006, Holford 2008, Dinley et al. 2010, Huckstorf and Wirkner 2011), functional morphology (Alba-Tercedor and Sánchez-Tocino 2011, Bond et al. 2008, Nickel et al. 2006, Patek et al. 2007, Wilhelm et al. 2011) and developmental studies (Postnov et al. 2002, Marxen et al. 2007, Puce et al. 2012) by studying species through a virtual, three-dimensional model. In palaeobiology, the

technique is, for example, frequently used to reveal the morphology and even anatomy of fossilised organisms that cannot be removed from their enclosure medium (Dierick et al. 2007, Dunlop et al. 2011, Hendrickx et al. 2006, Molineux et al. 2007, Penney et al. 2007, Sutton 2008). Most of these studies have imaged few or a single specimens, but some have harnessed the power of non-invasive, three-dimensional imaging to create vast amounts of data for large-scale systematic analyses (Wirkner and Prendini 2007, Ziegler et al. 2008, Golding et al. 2009, Zimmermann et al. 2011).

Despite the increasing use of these new imaging methods, most of the recently created datasets might not qualify for the notion of a cybertype. In most studies, specimens were prepared and imaged with a specific hypothesis in mind, focusing on certain morphological characteristics and omitting others, and the resulting data might thus not be useful for other purposes. Datasets that are intended to serve as a cybertype should fulfil at least the following three basic assumptions: (a) A cybertype should provide morphological and anatomical information of the same accuracy and reliability as provided by the physical type material, independently of a specific research question in mind; (b) A cybertype should be linked to the original type material, which can be consulted if in doubt. This implies that any method used to create the cybertype should not affect the morphological, anatomical and molecular identity of the original specimen (e.g. holotype, paratype or neotype); (c) A cybertype has to be retrievable and freely accessible. This involves making the data available through a reliable (internet) source under an open access licence and providing adequate security measures, such as archiving, backups and ensuring data format compatibility in the future, and allowing the annotation of the dataset with metadata in order to be retrievable and interpretable.

Towards this end, this study explores the potential of micro-computed tomography to create high-throughput morphological and anatomical data to support systematic and taxonomic studies by using polychaetes (bristle worms) as a demonstration taxon for macro-invertebrates. This taxon has been chosen because of the diversity of shapes and tissue types occurring among its members, allowing the investigation of the behaviour of the methodology across a range of samples with different characteristics. The outcomes are evaluated with regard to the first requirement for constituting a potential cybertype, that is, their ability to deliver reliable information on diagnostic and systematically important characters. However, from sample preparation to the final presentation of the results many steps are involved which may affect both the outcome of the data as well as the original specimen. Particularly, the imaging of soft tissues with micro-CT might require tissue staining (Metscher 2009a, b), but neither the effects of contrast-enhancing chemicals nor of ionising radiation upon the integrity of tissue and genetic material are yet fully understood. Although micro-CT radiation seems to negatively affect the genetic material of living tissue (Wolff 1971, Kersemans et al. 2011), no fragmentation of the DNA could be detected in museum material of bird skins exposed to micro-CT scanning (Paredes et al. 2012). The morphological and molecular integrity of scanned material is particularly important when valuable museum material is imaged, otherwise the material is rendered useless for further investigations. Therefore, by testing whether treatment with contrast agents or exposure

to X-ray radiation create structural damage to the tissue of the sample or impair the potential to amplify nucleic acid structures important for the molecular identification, this study assesses the compliance of micro-CT imaging with the second requirement for creating a cybertype. Finally, various aspects of exploring and communicating the resulting information through new ways of publishing are demonstrated and evaluated with regard to the third requirement for a cybertype. The paper concludes by summarising both the potential and the shortcomings of micro-CT imaging for taxonomic research and provides an outlook to possible future developments, including the overall applicability of the cybertype concept and the establishment of virtual collections.

Material and methods

Specimen preparation and processing

Nine polychaete specimens (seven different species) were chosen for this study, all of them in the clade Aciculata (Annelida, Polychaeta) (Table 1). Specimens are stored in the collections of the biodiversity laboratory of the Hellenic Centre for Marine Research, except for those of *Eunice roussaei* Quatrefages, 1866 (deposited in the Aristotelian University of Thessaloniki), *Alitta succinea* (Leuckart, 1847) and *Hermodice carunculata* (Pallas, 1776) (both subsequently used for molecular analyses and destroyed). All specimens had originally been fixed and preserved in different media, but were dehydrated to 96% ethanol prior to treatment. Identification was performed to the lowest possible level under a stereo microscope and light microscope, using the most recent literature available for each taxon (e.g. Böggemann 2002, San Martín 2003, Bakken 2004, Carrera-Parra 2006a, Zanol and Bettoso 2006). However, no dissections were performed, in order to assess whether internal characters required for identification in several species could be determined through virtual dissections instead.

To test the effect of different contrast-enhancement methods on the imaging results and tissue characteristics, several samples were treated with one of the following methods: a) tissue staining with 1% iodine in 96% ethanol; b) tissue staining with 0.3% phosphotungstic acid (PTA) in 70% ethanol; c) desiccation with Hexamethyldisilazane (HMDS). Protocols for both iodine and PTA staining follow Metscher (2009a). In both solutions, smaller samples were stained in 2 ml for 24 hours to several days, larger samples in PTA required longer staining (up to 3 weeks) in larger amounts (10 ml), the solution was renewed every five days to allow PTA to penetrate into the tissue. Samples treated with HMDS were left in the chemical for two to four hours, in the larger specimen (*Hermodice carunculata*) the chemical was renewed after two hours. The amount of HMDS and the treatment time depends on the size of the specimen: as a general guide, an amount twice the body volume of the specimen was used. Afterwards, specimens were removed from the chemical and left to dry for several hours, causing them to desiccate while retaining their morphology. Details on treatment for each specimen are presented in Table 1.

Table 1. Overview of scanned polychaete specimens, their preparatory treatment and scanning parameters. † = maximum dimension (in any direction) of the scanned part of the specimen; ‡ = multi-part scan, automatically joined; § = distance between projection images in degrees; | = total size of cross-section images in megabyte, file format: PNG/BMP

Species	Sample code	Location and depth	Size (mm) †	Scanned part	Sample preparation	Scanning medium	Rotation step §	Number of projection images	Exposure time per image (ms)	Total scanning time (h:min)	Resolution (µm per pixel)	Dataset size
<i>Lumbrineris latreilli</i> Audouin & Milne Edwards, 1834	ΦH20D	Gulf of Heraklion, Crete, Greece (35.3527°, 024.1084°), 20m	2.4	Anterior end	none	air	0.25°	1440	1925	3:08	1.919	506 (BMP)
<i>Eunice</i> sp. (juvenile)	CALA-10b-08	Alykes, Crete, Greece (35.4158°, 024.9875°), 10m	1.8	Whole animal	24 h in iodine	96% ethanol	0.26°	1384	1835	3:33	2.162	69 (BMP)
<i>Eunice</i> sp. (juvenile)	CELB-10c-08	Elounda, Crete, Greece (35.2516°, 025.7583°), 1m	2.2	Whole animal	24 h in PTA	96% ethanol	0.3°	1200	1835	3:05	2.298	93 (BMP)
<i>Eunice rousaei</i> Quatrefages, 1866	POL/EUN/503	Thermaikos Gulf, Greece	2.9	parapodium (mid-body)	2 h in HDMS	air	0.25°	1440	1155	2:19	2.878	326 (BMP)
<i>Alitta succinea</i> (Leuckart, 1847)	F6BNR1	Etang de Berre, France (43.4699°, 005.1699°), 2m	8.1	Anterior end	20 days in PTA, solution renewed every 5 days	96% ethanol	0.4°	900	190	2 x 0:17 ‡	3.425	1433 (PNG)
<i>Phyllodoce lineata</i> (Claparède, 1870)	Φ2H25E	Gulf of Heraklion, Crete, Greece (35.3640°, 025.1086°), 25m	3.9	Anterior end	36 h in PTA	96% ethanol	0.28°	1285	1165	2:05	1.645	450 (PNG)
<i>Phyllodoce</i> sp.	Φ2H40D	Gulf of Heraklion, Crete, Greece (35.3832°, 025.1086°), 40m	5.5	Anterior end and mid-body	36 h in PTA	96% ethanol	0.28°	1285	1336	2:04	1.096	2406 (PNG)
<i>Syllis gracilis</i> Grube, 1840	CELB-1a-07	Elounda, Crete, Greece (35.2516°, 025.7583°), 1m	7.6	Anterior end and mid-body	3 days in iodine	96% ethanol	0.25°	1440	1165	3 x 1:52 ‡	1.439	2969 (PNG)
<i>Hermiodice carunculata</i> (Pallas, 1776)	–	Alykes, Crete, Greece (35.4158°, 024.9875°), 5m	10.1	Anterior end	2 x 2 h in HMDS	air	0.3°	1200	155	0:32	8.922	877 (BMP)

Wet samples were scanned in heat-sealed 200 µl polypropylene pipette tips, either in ethanol or in air. The top of the container was sealed with a plasticine cap to prevent the specimen from drying out during scanning (for a similar setup see Metscher 2009a). Samples dried with HMDS were partially enclosed in a small piece of styro-foam which in turn was mounted on a thin metallic sample holder. For assessing the quality of the scans with regard to distinguishing features, in this study only the anterior end of most worms was scanned. Scanning only the anterior end reduced scanning time and allowed us to choose a higher resolution. In polychaetes, the anterior end usually comprises most diagnostic characters, thus allowing us to assess the usefulness of the scans based on taxonomic criteria.

Image acquisition

Samples were imaged with a SkyScan 1172 microtomograph (<http://www.skyscan.be/products/1172.htm>) at the biodiversity laboratory of the Hellenic Centre for Marine Research. This system uses a tungsten source with energies ranging from 20–100kV and is equipped with an 11 megapixel CCD camera (4000×2672 pixel) with a maximal resolution of <0.8 µm/pixel. Specimens were scanned at a voltage of 60 kV with a flux of 167µA and scans were performed for a full rotation of 360°. Except for *Alitta succinea*, for which a camera pixel binning of 2 × 2 was chosen, images were always acquired at highest camera resolution. Individual scanning parameters can be found in Table 1. Projection images acquired during the scanning process were subsequently reconstructed into cross sections with SkyScan's NRecon software which employs a modified Feldkamp's back-projection algorithm. Sections were always reconstructed from the total number of projection images (360°) to obtain a greater level of detail, other reconstruction parameters were chosen individually for each sample. In case of strong density differences in the scanned sample, the upper limit of the grey scale histogram was lowered to unite very dense values. This causes dense values above the set limit to be assigned to the same grey scale value without differentiation and allows softer (less dense) tissues to be visualised with greater detail. The lower limit of the histogram was set at the value for the surrounding medium (air or ethanol). To reduce the size of the resulting images, only areas containing relevant data (regions of interest) were reconstructed, thus excluding the surrounding air or enclosure medium.

Molecular analyses

Hediste diversicolor (O.F. Müller, 1776) specimens collected in Tsopeli lagoon in Amvrakikos Gulf (Western Greece) were sequenced before and after X-ray exposure in order to assess whether the radiation had an effect on the 16S rRNA sequence obtained. Samples were exposed either to high energy of radiation for a relatively short time (100kV for 1.5h) or repetitively exposed for three cycles of 12 hours at medium

energy (12h, 24h, 36h at 60kV). In the latter series, some tissue was removed from the specimen for DNA extraction after each cycle. A fragment of the 16S rRNA gene (~500 bp) was amplified using a primer pair designed for polychaetes: 16SAN-F (TACCTTTTCATCATGG) and 16SEU-R (ACCTTTGCACGGTCAGGRTACCGC) (Zanol et al. 2010). Genomic DNA concentration of samples was calculated using a Nanodrop 1000 spectrophotometer (Wilmington DE, USA) on wavelength 260 and 280nm. Polymerase Chain Reactions (PCRs) were performed in a final volume of 20 µl containing 0.4 units KAPA Taq DNA polymerase (Kapa Biosystems, Inc, USA), 2 µl PCR buffer 10x, 4.5 mM MgCl₂, 2 mM dNTPs, 10 pmol of each primer, and 0.5 µl (~5 ng/µl) of DNA template. Amplification was performed in a MJ Research PTC-200 Thermo Cycler (Harlow Scientific, USA) programmed as follows: initial denaturation cycle at 96°C (4 min) followed by 35 cycles of denaturation at 93°C (45 sec), annealing at 52°C (1 min) and extension at 72°C (1 min); last cycle was followed by a final extension at 72°C for 7 min. The sequences were processed with MEGA v. 5 software (Tamura et al. 2011). Obtained sequences were submitted to GenBank (Benson et al. 2005) under the accession numbers KC113440-KC113445.

Processing, presentation and dissemination of image data

Two-dimensional images:

All resulting datasets of cross sections were post-processed with the CTAnalyzer (CTAn) software (SkyScan, Kontich, Belgium) by selecting a Region of Interest (ROI) containing the sample but removing further superfluous information, thus creating a dataset of reduced size. To obtain a three-dimensional representation of the sequence of cross section images, the data were visualised with two different volume rendering software packages: both CTVox (SkyScan, Kontich, Belgium) and the free software Drishti were employed (<http://anusf.anu.edu.au/Vizlab/drishti/>). Volume rendering displays the data by assigning a colour value and an opacity value to each data point (voxel) in the dataset. By changing these transfer values, different features of the dataset can be visualised and explored. Density-based false-colour renderings were applied to the data where this was considered helpful to visualise structures. Isosurface models (geometrical representations of surfaces of equal values) were created with Amira v. 5.2 (Visage Imaging, Berlin, Germany). Two-dimensional images were extracted as bit-map files with the image export function of the respective software and consequently cropped to final dimensions and minimally edited in Adobe Photoshop to enhance contrast (adjusting image levels and curves) or transform colour tint (adjusting hue and saturation), as well as to add annotations.

Interactive volumetric data:

The *Lumbrineris latreilli* dataset was first processed with custom functions of CTAn (thresholding, smoothing, noise removal) to isolate the jaws from the surrounding tissue and saved as a separate dataset. This new dataset was subsequently loaded into

the free image editor Fiji (<http://fiji.sc>) and reduced in size to a stack of 320 images with dimensions of 205×173 pixels. These bitmaps were converted into TGA (Truevision Graphics Adapter) files with the free ImageMagick tool <http://www.imagemagick.org> and rendered with the C++ programme volren (Ruthensteiner et al. 2010) which is based on the plotting library S2PLOT (Barnes et al. 2006). With this library, the data were converted into a three-dimensional VRML (Virtual Reality Modelling Language) object, accompanied by a PNG (Portable Network Graphics) file of each angular view. A corresponding script provided by the authors ensures that the correct view is rendered when the object is manipulated. In the volren script parameters were adjusted to AMIN=0.0001 and AMAX=0.1 and the colour map “iron” was chosen. In the resulting VRML file the texture transparency parameter was changed from 0.4 to 0.1 throughout the file in order to increase the contrast of the embedded model in the PDF (Portable Document Format) file.

Interactive surface description models:

Using the segmentation editor of Amira, features of interest were manually segmented (“labelled”) with the brush tool. For each feature a new LabelField was created, thus allowing the different objects to be manipulated separately at later stages. Labelled features were converted into surfaces with the SurfaceGen module and where the number of polygons was too high (>1,000,000) they were reduced with the Simplifier tool to increase computation efficiency during further processing. Amira’s SmoothSurface module did not produce satisfactory results in models with small detailed structures, since the module does not allow for selective smoothing and small structures disappeared after the application of the module. The models were, therefore, exported as OBJ (Wavefront Object) files and further processed with Blender 2.63a (<http://www.blender.org>), a high end, open source, 3D design program. In Blender, the surface was cleaned of artefacts by applying the Vertices’ Relaxation and the Vertices Smoothing operations, which replaced the original model’s points (vertices) in average positions between them, thus smoothing surface anomalies. This process was repeated where necessary, until the model’s surface appeared smooth and even, without alienating main parts or the overall morphology of the model. If surface noise still persisted in parts of the model, then a second, manual part of cleaning was applied. In this case, specific parts of the model were selected individually and corrected by using additional tools in Blender (e.g. sculpting smooth brush). Some geometrically elegant (small, narrow, light) parts of the model (e.g. chaetae) could not be cleaned or were destroyed by the above techniques, in this case these parts and elements were reconstructed (remodelled) manually and added again to the model.

Videos:

The videos of 3D volume renderings were created with CTVOx, using the flight recorder function, and saved as an AVI (Audio Video Interface) file. The video of cross sections through the sample was created by loading the image stack of the sections into Fiji and re-sampling the data to reduce image size. The resulting new stack was exported as an AVI video file with 20 frames per second. To be able to embed the vid-

eos into the PDF document, the AVI files were converted with an online conversion software (<http://www.online-convert.com>) into FLV (Flash Video) files with a bit rate of 1000kb/s and a width of 400px.

Embedding multimedia and interactive objects into PDFs:

Both surface models and volume renderings were embedded into the PDF with the Acrobat X Pro 3D PDF Converter Suite (Tetra 4D, Seattle, USA). The 3D Reviewer module was used to define colours and views and to add annotations. The resulting data were again exported to PDF format, specifying in the export options “PCR tessellation” and “Compress tessellation”. These options reduced the final file size to about one third of the original object size. The proper rendering of the volumetric data required also the inclusion of a JavaScript file which is distributed along with the S2PLOT library (<http://astronomy.swin.edu.au/s2plot/peripheral/s2plot.js>). Videos were added with Adobe’s Add Multimedia function. The process of embedding models and multimedia content into PDFs is relatively straightforward, detailed descriptions are provided by Barnes et al. (2006), Ruthensteiner and Hess (2008), Kumar et al. (2010) and Ruthensteiner et al. (2010).

Electronic publication and data dissemination:

All media included in this publication as well as supporting material (surface models, image files, videos) are published under a Creative Commons Attribution 3.0 (CC-BY) licence in a Virtual Research Environment, the Polychaete Scratchpad (<http://polychaetes.marbigen.org>). The full volumetric datasets have been archived at the Dryad Data Repository (<http://datadryad.org>, doi: 10.5061/dryad.84m54). Since the direct inclusion of interactive, three-dimensional models in web pages is still in its infancy and requires specific browser and driver configurations on the client side, interactive models have been included as separate PDF files on the web site of the journal as well as on the Scratchpad site, thus allowing the majority of users to access this content. Most Acrobat products (Reader, Professional) from Version 8 onwards support the display of embedded media. However, specific versions of the software still show incompatibility problems and some users might therefore encounter problems viewing the interactive content. In this case, it is recommended to download the multimedia content (videos, interactive models) from Dryad, or the Polychaete Scratchpad, and view it with other software (e.g. a multimedia player).

Results**Information content of the datasets*****Lumbrineris latreilli*:**

Only the anterior end of the specimen was scanned, the total length of the visualised part being 2.4 mm. At a resolution of 1.9 $\mu\text{m}/\text{pixel}$, the smallest discernible structures

are about $4\ \mu\text{m}$ in size. Since the specimen has been scanned in air without prior desiccation, a thin film of ethanol partially covers the body and obscures parts of the external morphology; however, general external characters (size and shape of larger features such as segments, head, mouth opening) are clearly recognisable (Fig. 1a). The shape of the parapodia is visible but obscured by the ethanol film. The number and arrangement of the chaetae can be discerned but the chosen resolution prevents finer details such as chaetal articulation or dentation from being captured. The internal anatomy, on the contrary, has been recorded in substantial detail. The resolution is high enough to allow the observation of even the fine structure of the vascular system in the anterior part of the head (Figs 1b–c). Likewise, muscular groups and even their fibres are clearly visible (Fig. 1d). Nervous tissues (brain, ganglia, nerves) have a very low density and are difficult to depict in detail in the volume renderings. However, large nervous

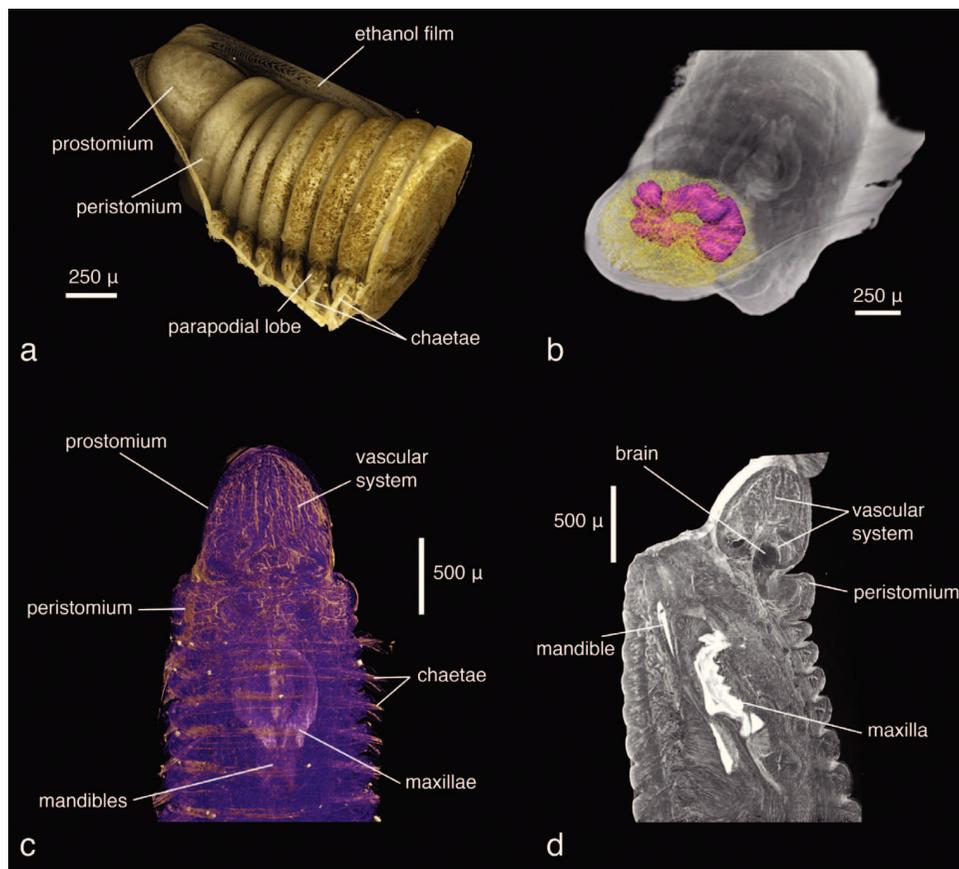


Figure 1. *Lumbrineris latreilli*, **a** false-colour volume rendering of the anterior region, dorso-lateral view, specimen partly covered by a thin film of ethanol **b** surface model of the vascular system (yellow) and brain (purple) superimposed on volume rendering of the worm, dorso-anterior view **c** false-colour, semi-transparent volume rendering of the anterior region, dorsal view **d** volume rendering of virtually dissected anterior region, lateral view.



Video 1. *Lumbrineris latreilli*, false-colour volume rendering, virtual dissection of anterior end, dorsal view. Video available for download in full resolution from <http://polychaetes.marbigen.org/lumbrineris-latreilli-micro-ct-video>.

structures can be visualised by applying appropriate transfer functions (Video 1) and through careful remodelling (Fig. 1b). The most prominent visible feature is the jaw apparatus, a calcified complex system of maxillae and mandibles (Figs 2–4). However, the maxillary pair MV, the maxillary carriers, as well as the accessory lamellae of mandibles III and IV appear to have a similar density to the surrounding muscle tissue. This makes it impossible to visualise them in the low-resolution interactive volume model included in this publication (Fig. 3). In the high-resolution dataset, however, they can be visualised by applying suitable transfer functions (Figs 4b–c), which allows them to be included in the surface model (Fig. 2). The accessory lamellae of MI and the connecting plates of MI and MII, described as “weakly sclerotised in *Lumbrineris*” by Carrera-Parra (2006b), cannot not be clearly discerned in the data.

***Eunice* sp. (juveniles):**

The specimens were similar in size and developmental stage, making them ideal for testing the effect of different tissue stains (discussed below). As in the *Lumbrineris* dataset, larger external features such as body shape, antennae and parapodia are excel-

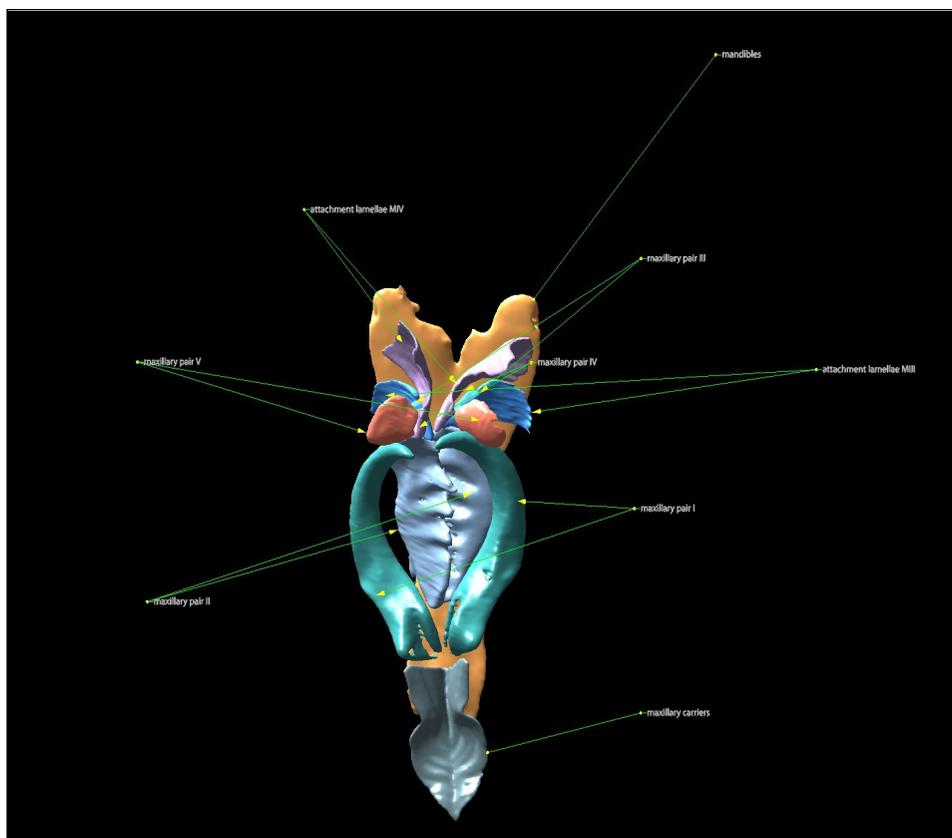


Figure 2. *Lumbrineris latreilli*, surface model of the jaw apparatus. Terminology follows Carrera-Parra (2006). If viewed with Adobe Acrobat Reader (version 8 or higher), the interactive 3D-mode can be activated by clicking on the image, allowing the user to rotate, move and magnify the model, to isolate elements and to change the light settings.

lently visible (Figs 5a–b), although the animals are smaller and the resolution is slightly coarser (ca 2.25 $\mu\text{m}/\text{pixel}$). No details of chaetae are visible, but the subacicular hooks are clearly depicted especially in the iodine-stained specimen (Fig. 5b) and the chaetal bundle is well visible in cross sections (Fig. 6). Internally, large organs and muscle groups are fairly distinct (Fig. 6), but the small size of the animal (width ca 0.3 mm) does not allow details that would be visible in histological sections to be discerned (e.g. vascular system, ganglia), with the exception of muscle fibres. The general structure of the jaw apparatus is evident, but finer details are difficult to see, and only MI–MIII can be unambiguously identified (Fig. 5c).

***Eunice roussaei*:**

The dissected parapodium shows, especially in the cross sections, the exact arrangement of the different chaetal types (aciculae, subacicular hook, compound chaetae, supra-ac-

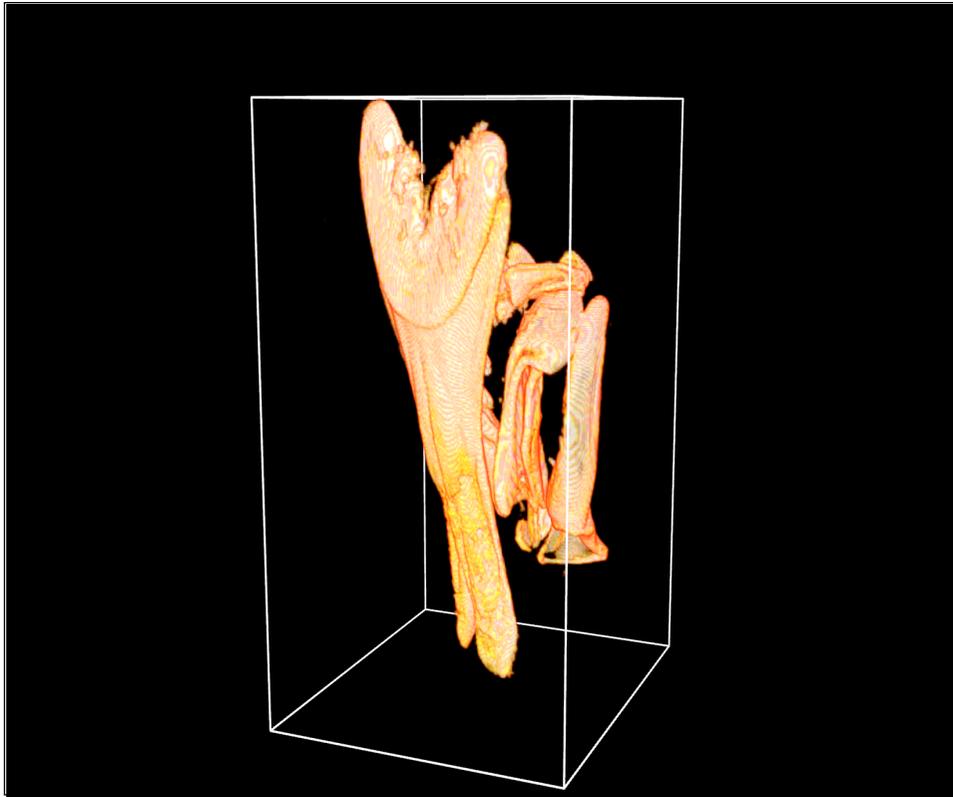


Figure 3. *Lumbrineris latreilli*, volume model of the jaw apparatus. Note the absence of MV as well as the accessory lamellae of MI, MIII and MIV which cannot be displayed in this low-resolution version. If viewed with Adobe Acrobat Reader (version 8 or higher), the interactive 3D-mode can be activated by clicking on the image, allowing the user to rotate, move and magnify the model.

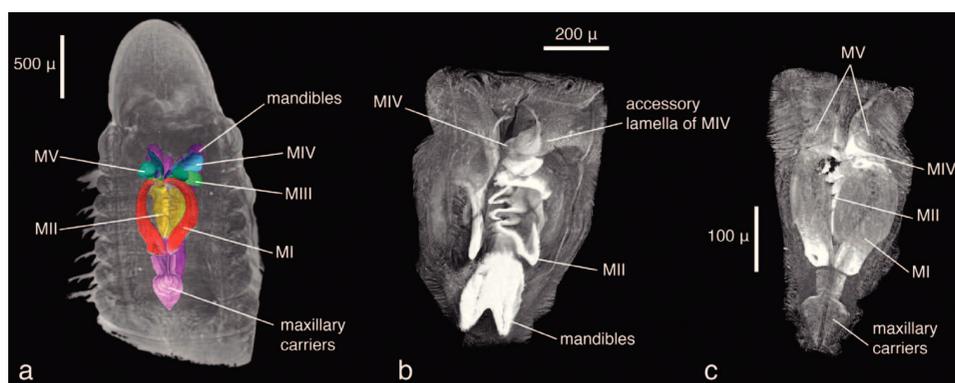


Figure 4. *Lumbrineris latreilli*, volume rendering of **a** anterior region with maxillary apparatus superimposed as a surface model, dorsal view **b** virtually dissected jaw apparatus, ventral view with mandibles partly removed **c** same, dorsal view.

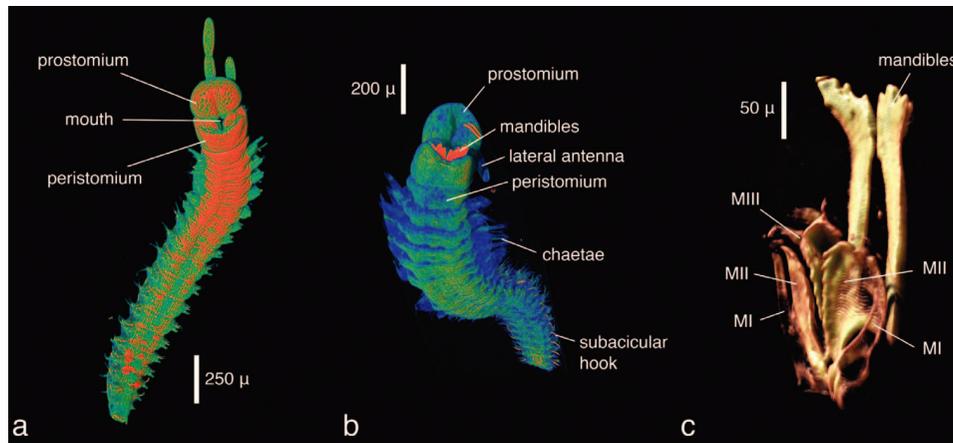


Figure 5. *Eunice* sp., false-colour volume renderings (colours indicate relative densities: blue: low, green: medium, red: high) of **a** iodine-stained specimen, ventral view **b** PTA-stained specimen, ventral view **c** false-colour volume rendering of virtually dissected maxillary apparatus (iodine-stained individual), dorso-lateral view.

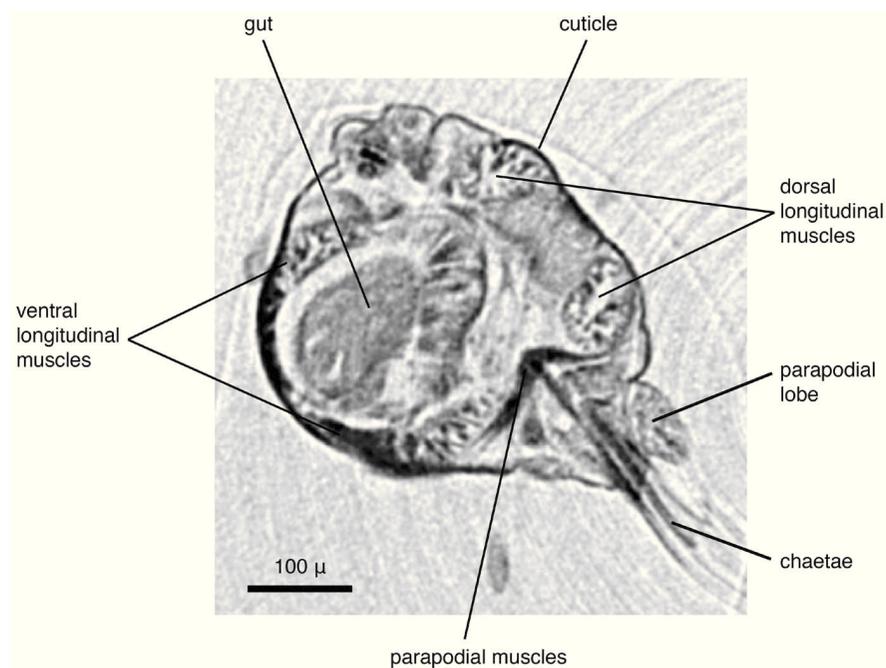


Figure 6. *Eunice* sp., transaxial cross section at mid-body (iodine-stained individual).

icular limbate and pectinate chaetae as well as internal notopodial chaetae) (Figs 7a–b). External features, such as dorsal and ventral cirri, parapodial lobes, branchial stem and branchial filaments, can likewise be observed (Fig. 7c). Due to the position in which the specimen dried, the branchial stem is slightly recoiled, making the exact count of

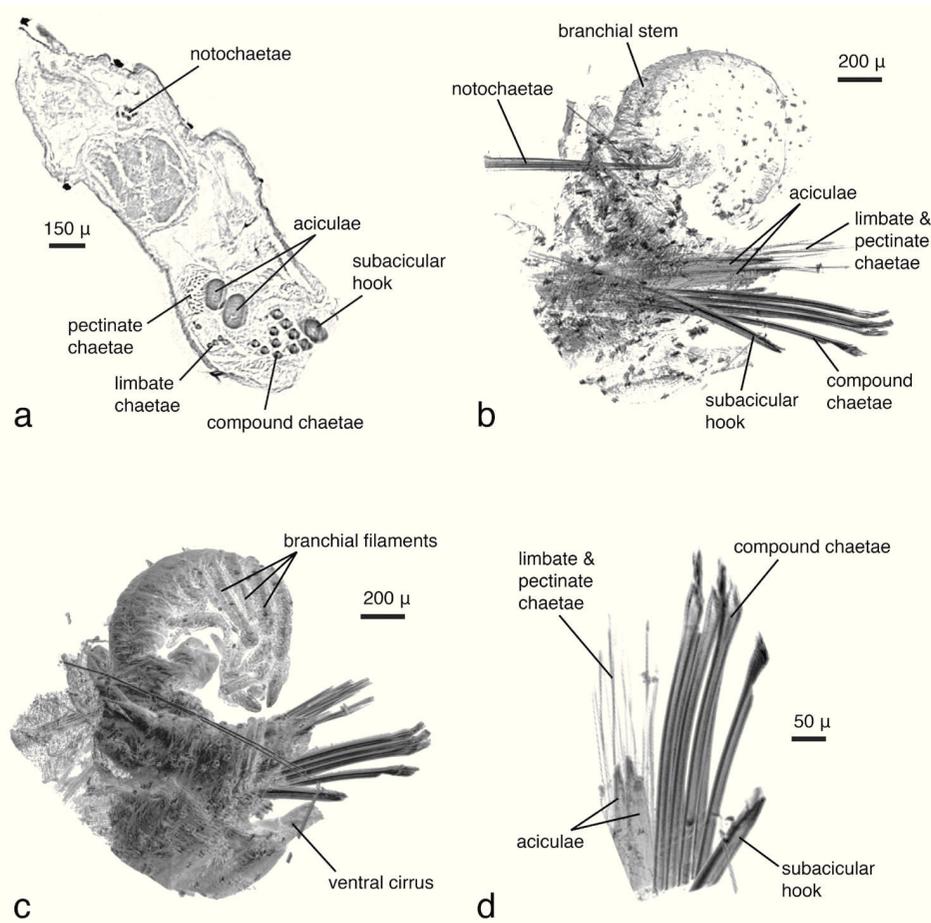


Figure 7. *Eunicie roussaei*, mid-body parapodium, **a** cross section through parapodial base **b** semi-transparent volume rendering **c** opaque volume rendering **d** volume rendering of chaetae.

branchial filaments not straightforward; they are better delineated in the interactive, three-dimensional model of the specimen (Fig. 8). Details of chaetae (e.g. serration, dentation) are not visible (Fig. 7d). This is primarily caused by the coarse resolution of the scan (ca 2.9 $\mu\text{m}/\text{pixel}$), which prevents capture of these tiny details. Furthermore, because the specimen was dried and scanned in air, the chaetae vibrate slightly during each rotation step, resulting in a slight blur in the final images (Fig. 7d). Several internal structures are visible, such as muscle groups or connective tissues, but no blood vessels (e.g. in the branchiae) can be seen. Generally, most internal structures are difficult to identify, since the contrast between neighbouring tissues was equalised by the desiccation process.

Alitta succinea:

The overall morphology of the specimen, especially cephalic features (shape, appendages) and parapodia, can be clearly observed. The true three-dimensional structure of

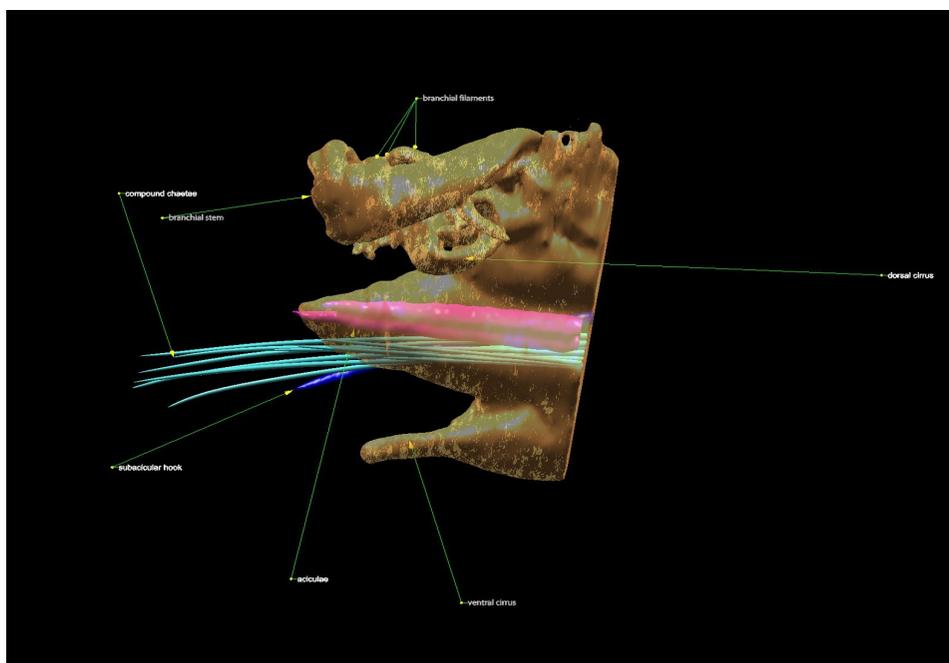


Figure 8. *Eunice roussaei*, surface model of mid-body parapodium. Limbate chaetae, pectinate chaetae and notochaetae are not shown. Compound chaetae have been included as simplified, remodelled version and do not depict true shapes. If viewed with Adobe Acrobat Reader (version 8 or higher), the interactive 3D-mode can be activated by clicking on the image, allowing the user to rotate, move and magnify the model, to isolate elements or to change the light settings.

the parapodia, which consist of a complex arrangement of parapodial lobes, as well as dorsal and ventral cirri, are best distinguished in stereo display (e.g. with anaglyph 3D glasses) (Figs 9a–b, visible with red-cyan glasses). Chaetal structures cannot be observed, specimen preparation as well as scanning parameters were inappropriate for their proper visualisation. The pharynx in this specimen is everted, thus the determination of the shape of the jaws as well as of the paragnath shapes (conical) and distribution patterns—important taxonomic characters—is straightforward without having to virtually dissect the specimen (Fig. 9a). Since X-ray imaging is attenuation-based and thus records differences in density (or, depending on the energy, differences in the atomic number of the material), the colour of the paragnaths (another diagnostic character) cannot be observed. Eyes are visible as slightly darker spots on the epidermis, their density is apparently slightly different from the surrounding tissue (Fig. 9d). The internal anatomy is, as in the other samples, well visible. Major muscle groups can be discerned (Fig. 9c, e), as well as the vascular system. The latter seems to be discontinuous in parts (Fig. 9c), possibly a result of fixation in ethanol. The brain or other nervous tissues cannot be discerned. Apparently hollow regions in the pharyngeal area result from insufficient staining (Fig. 9f). The density of these unstained regions is too low to be visualised and the corresponding information is lost.

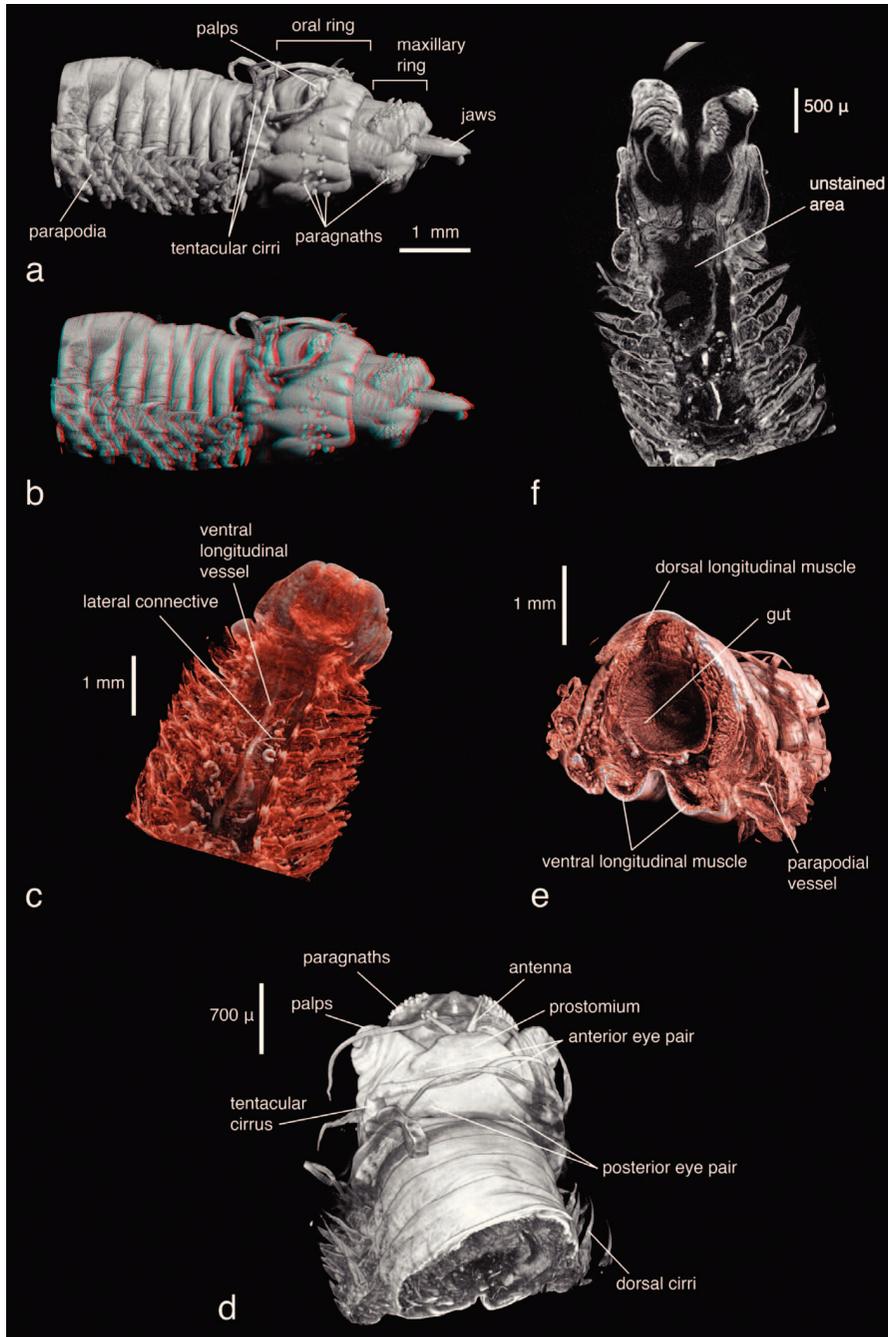


Figure 9. *Alitta succinea*, **a** normal view and **b** stereo view of volume-rendering of anterior region, lateral view (3D effect is revealed when viewed with red-cyan glasses) **c** coronal cross section of anterior region depicting areas of insufficient tissue staining with PTA. Black areas in pharyngeal region remained unstained and information is lost **d** dorsal view and **e** posterior-lateral view of false-colour volume rendering of virtually dissected anterior region **f** volume rendering of anterior region, dorsal view.

***Phyllodoce lineata* and *Phyllodoce* sp.:**

Phyllodoce lineata can be identified to species level almost solely on the basis of the virtual specimen; all taxonomic key characters are visible, with the exception of mid-body parapodia which were not included in the imaged part of the specimen. However, the number, shape, length and arrangement of tentacular cirri and antennae, the shape (fused, partly fused, covered) of the prostomium and the anterior segments of the parapodia with dorsal and ventral cirri, and the position of chaetae can be observed (Fig. 10a), but finer details of the parapodial lobes and the chaetae cannot. The pharynx is partly everted, which makes the assessment of the arrangement of the pharyngeal papillae straightforward. The subdivision of the pharynx is well visible, with the proximal end being covered with scattered smaller papillae, the distal end with six rows of large papillae (Fig. 10a). The papillae around the pharyngeal opening can only be seen when the animal is virtually dissected (Fig. 10b). The eyes, containing lenses as in most phyllodocid species (Rouse and Pleijel 2001), appear as dense structures (Figs 10c–e, Video 2). Internally, all muscular features, the gastrointestinal tract, ganglia and the



Video 2. *Phyllodoce lineata*, false-colour volume rendering, virtual dissection. Video available for download in full resolution from <http://polychaetes.marbigen.org/phyllodoce-lineata-micro-ct-video>.

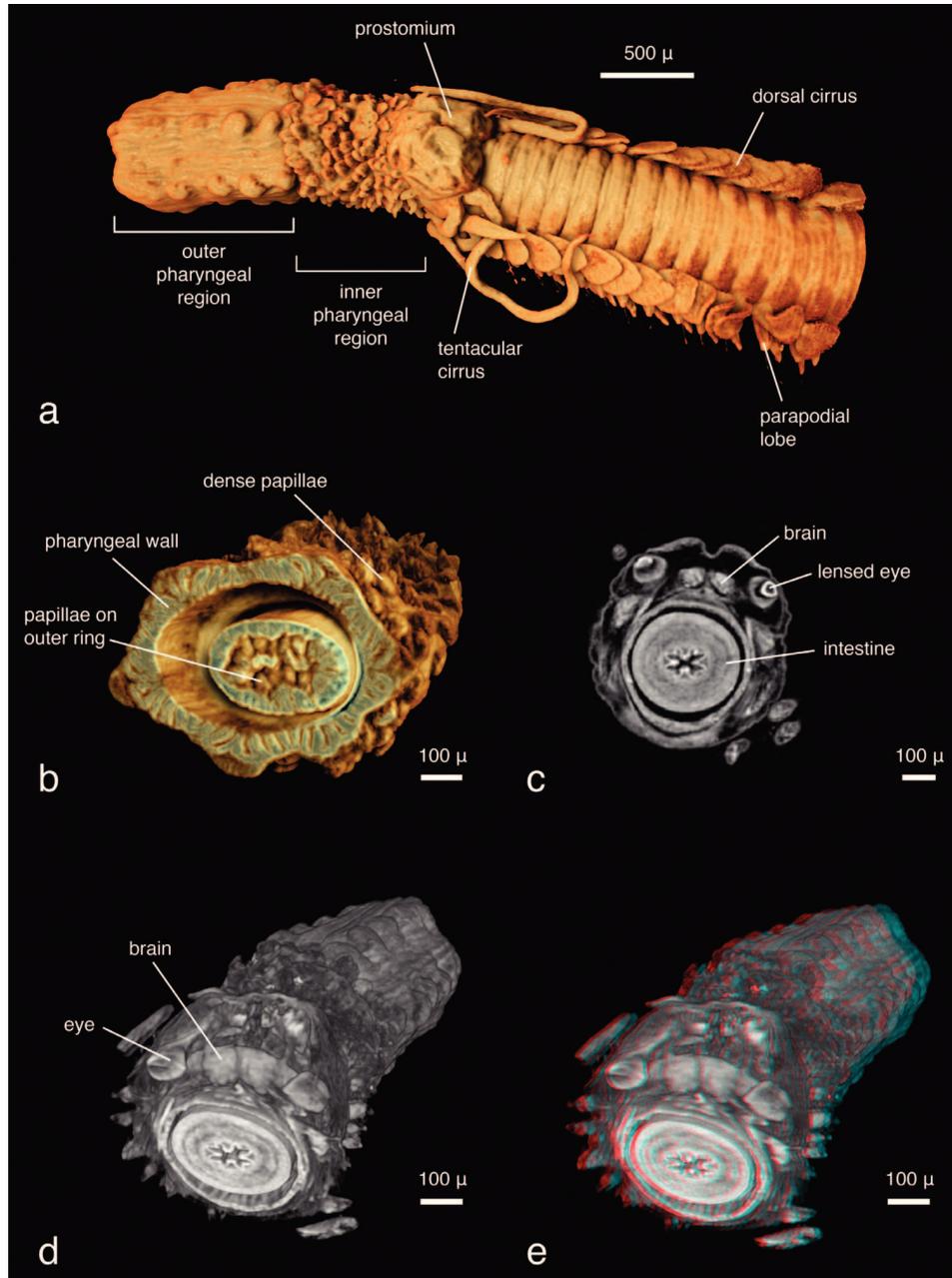


Figure 10. *Phyllodoce lineata*, **a** false-colour volume rendering of anterior region, dorso-lateral view **b** false-colour volume rendering of virtually dissected distal end of partly everted pharynx, focusing on terminal pharyngeal papillae, anterior view **c** transaxial cross section at eye level **d** volume rendering of virtually dissected anterior region showing the brain, dorso-posterior view **e** same, in stereo view (3D effect is revealed when viewed with red-cyan glasses).

brain and large connecting nerves are well visible (Figs 10d–e, 11a, Video 2), especially when viewed in stereo display with red-cyan glasses (Fig. 10e). The other *Phyllodoce* specimen was imaged with a similar resolution and can thus be compared to the scan of *P. lineata*. Morphologically, no differences between the two specimens are evident, however, the pharynx in *Phyllodoce* sp. is not everted. Virtual dissections reveal two subdivisions of the pharynx (Fig. 11b), a distal part covered with large papillae and a proximal part with small papillae. The cross sections of the pharyngeal regions show that the distal part has six rows of large papillae (Fig. 11c), however, neither the number nor arrangement of the smaller papillae can be determined (Figs 11d), making an unambiguous identification of the species impossible. Naturally, no colour patterns — a species-specific character in several phyllodocid taxa (Nygren et al. 2010, Nygren and Pleijel 2011) — are visible in any of the scans.

Syllis gracilis:

As in the other data, large external morphological features of the specimen are well defined. Appendages, their articulation and number of articles are clearly depicted (Fig. 12a). Internally, features such as muscles (Figs 12b, d), the brain (Fig. 12c), the gastrointestinal tract (Fig. 12d) and muscle groups with their individual fibres can be identified. Other features such as eyes, pharyngeal papillae and the pharyngeal tooth are difficult to detect in volume renderings but can be observed in cross sections (Figs 13a–c).

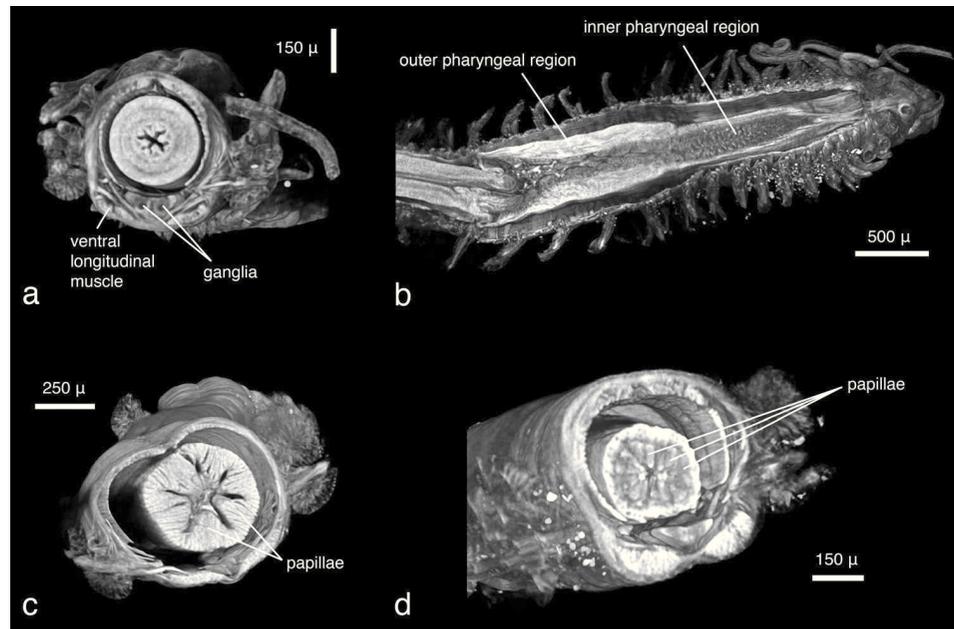


Figure 11. **a** *Phyllodoce lineata*, volume rendering of **a** anterior end, virtually dissected behind prostomial region, posterior view **b** *Phyllodoce* sp., volume rendering of virtually dissected pharynx, dorsal view **c** *Phyllodoce* sp., virtual dissection of distal pharyngeal subdivision with large papillae, anterior view **d** *Phyllodoce* sp., virtual dissection of proximal pharyngeal subdivision with small papillae, antero-lateral view.

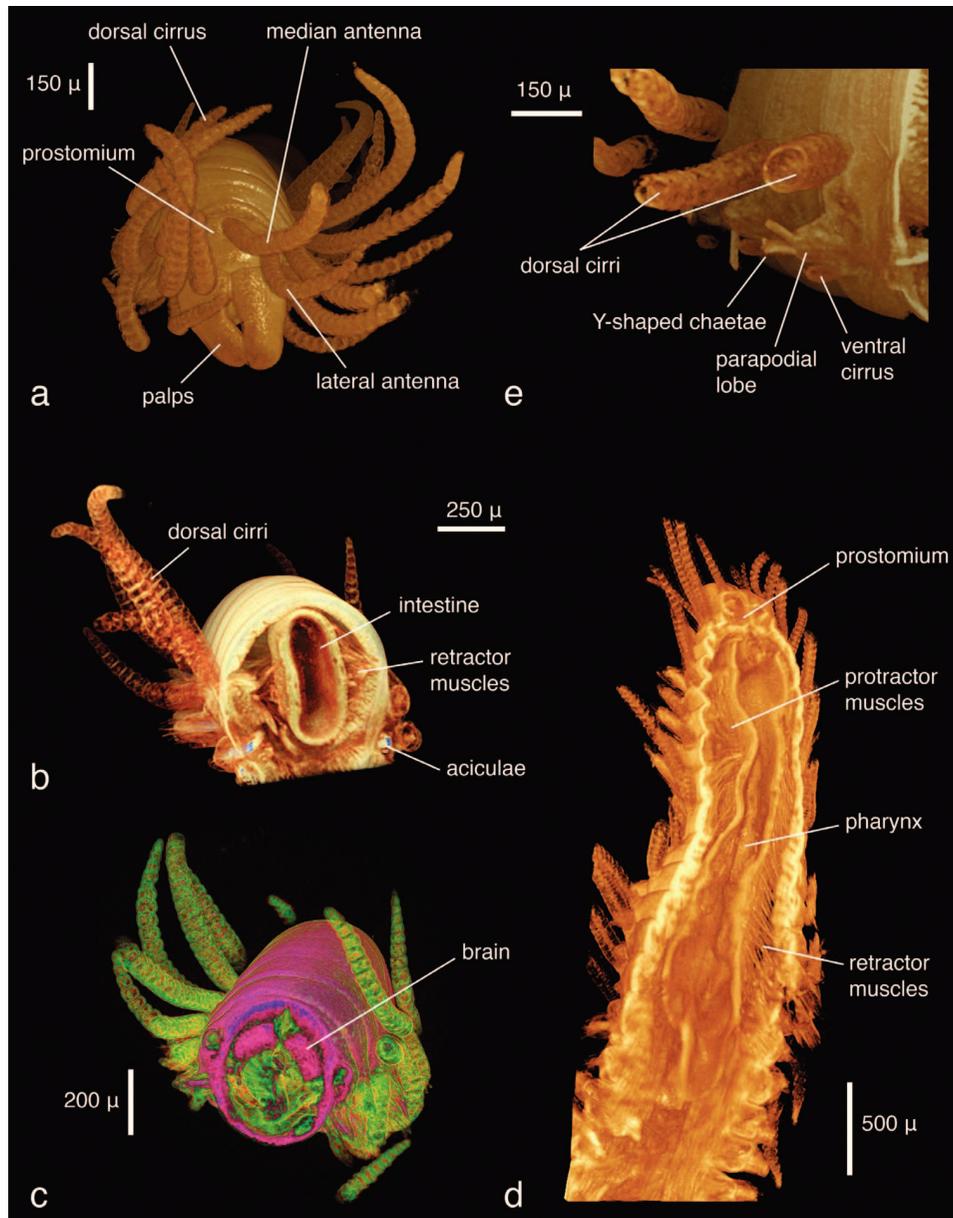


Figure 12. *Syllis gracilis*, false-colour volume rendering of **a** anterior region, anterior view **b** anterior region virtually dissected at pharyngeal level, posterior view **c** virtually dissected anterior region showing the brain, anterior view **d** virtually dissected region in front of proventricle, dorsal view; **e**) mid-body parapodia, posterior-lateral view.

The length of the proventricle and the number of its muscle rows are likewise important diagnostic systematic characters in the family. The strong muscle fibres of the proventricle are extremely well visible, and through virtual dissection their three-dimensional

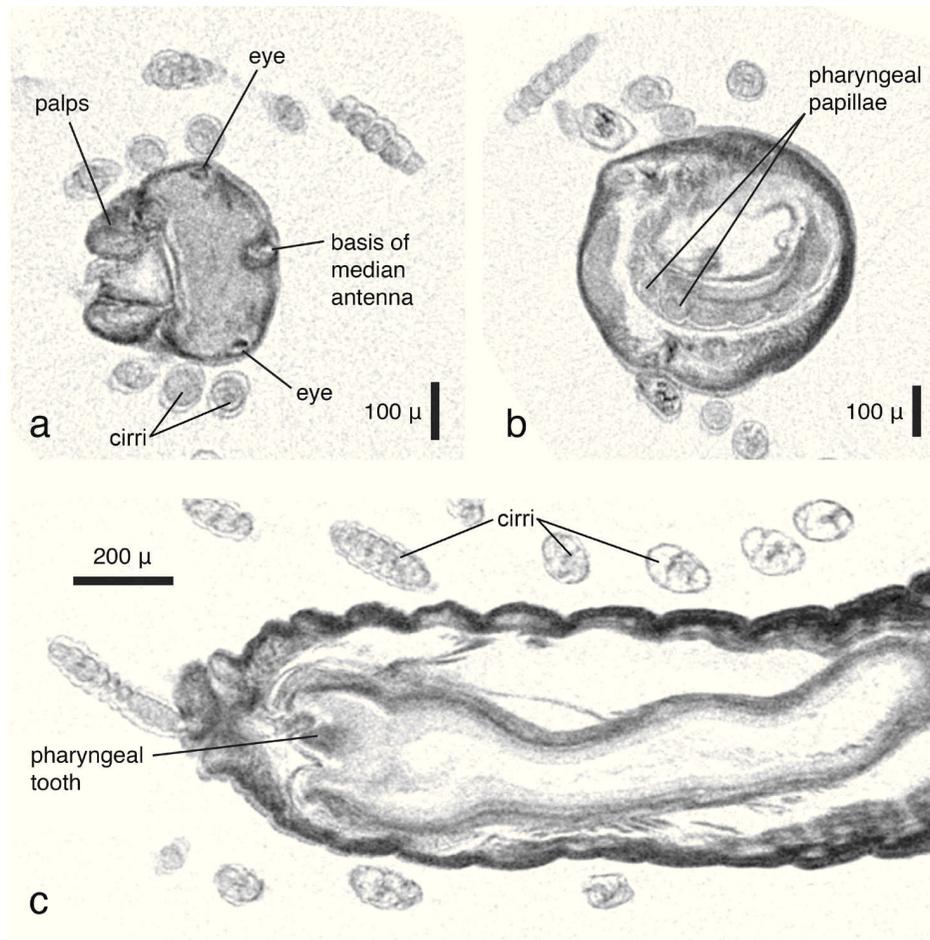


Figure 13. *Syllis gracilis*, cross sections, **a** coronal view showing prostomium, palps and posterior eye pair **b** transaxial view at level of pharyngeal opening, showing papillae around pharyngeal opening **c** dorsal view, pharyngeal opening and pharyngeal tooth. Double lines at borders of cirri are artefacts resulting from either movement of specimen during the scan or from settings during dataset reconstruction.

arrangement as well as their number can be well observed (Fig. 14). However, one of the key characters in syllid systematics are fine differences in chaetal structures. These cannot be discerned with the present resolution; only the rather large and robust Y-shaped chaetae typical for *Syllis gracilis* are visible (Fig. 12e).

***Hermodice carunculata*:**

This specimen has been scanned at a low resolution (8.9 $\mu\text{m}/\text{pixel}$) because of its large size, thus details of smaller structures such as chaetae are lost. Individual chaetae can be differentiated, but at the base of the parapodia they appear as a thick, merged structure (Figs 15a, d). Larger external morphological features such as branchiae, parapodial structures, antennae and the caruncle are all clearly visible (Fig. 15a). Eyes,

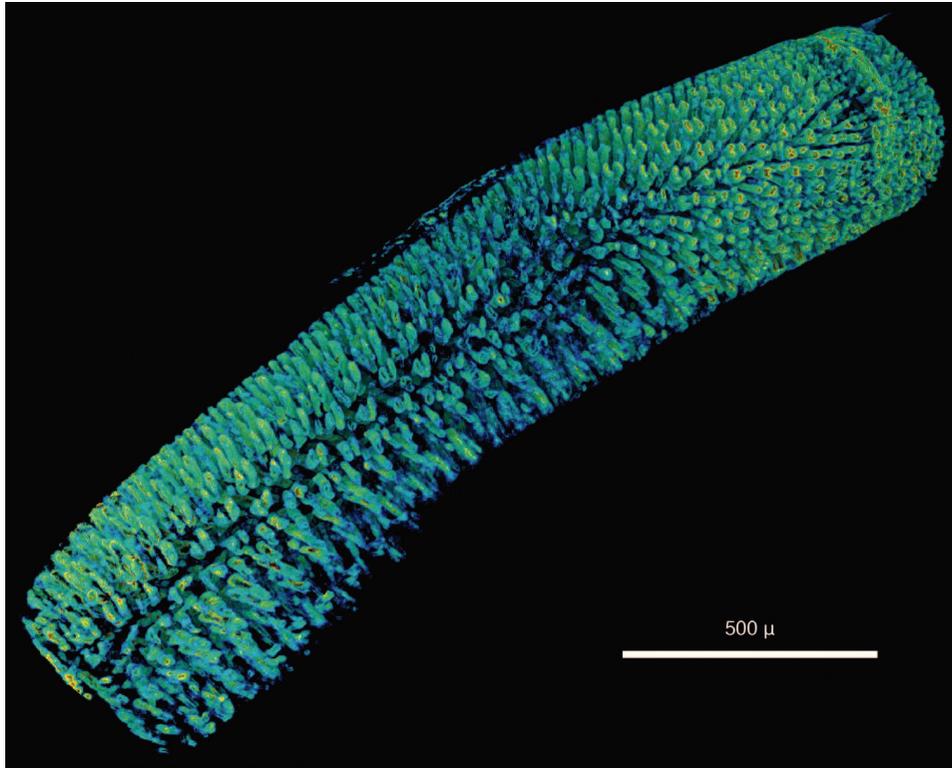


Figure 14. *Syllis gracilis*, false-colour volume rendering of virtually dissected proventricle.

located under the cuticle, are visible in virtual dissection (Fig. 15b). Internally, the heavily vascularised area around the pharynx is well defined (Fig. 15c, Video 3). The complex folds of the pharyngeal system can be observed (Fig. 15b, Videos 3–4), as well as various parts of the muscular system (pharyngeal muscles, longitudinal muscles, parapodial muscles). The brain is clearly visible (Fig. 15b) and large nervous fibres such as the circumesophageal connective can be traced from the brain to the ventral ganglia (Fig. 15d).

Contrast enhancement

Two different contrast enhancement techniques were employed in this study. These involved the removal of the surrounding liquid medium and tissue staining with electron-dense substances. Both result in an increased density difference between the specimen and the surrounding medium and thus produce sharp, contrasting images. However, different methods accentuate different morphological and anatomical features and thus the information content of the data differs accordingly. *Lumbrineris latreilli* was simply scanned in a sealed tube to prevent the specimen from drying out

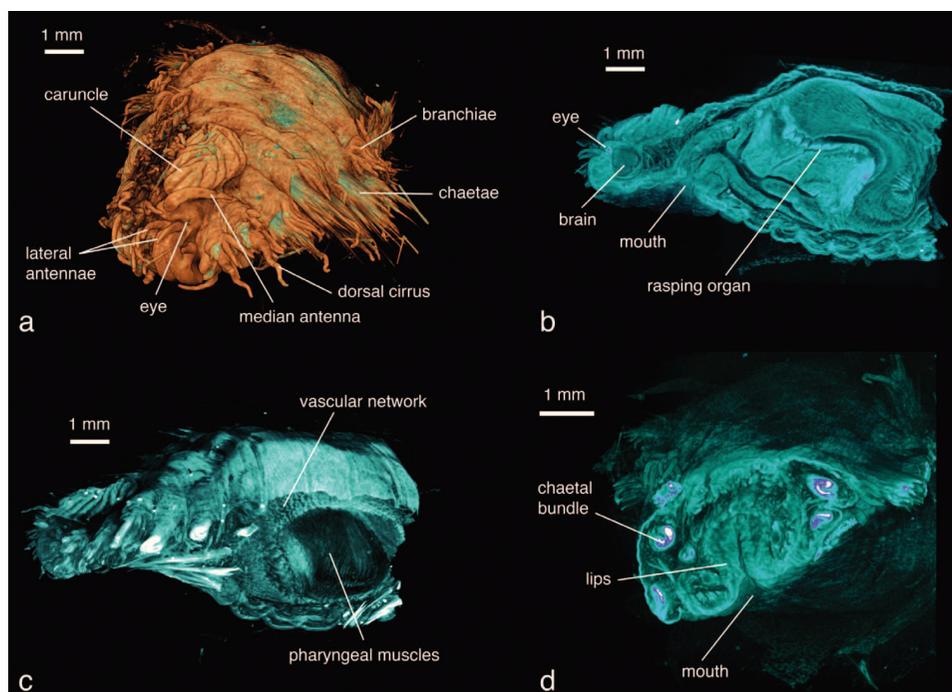
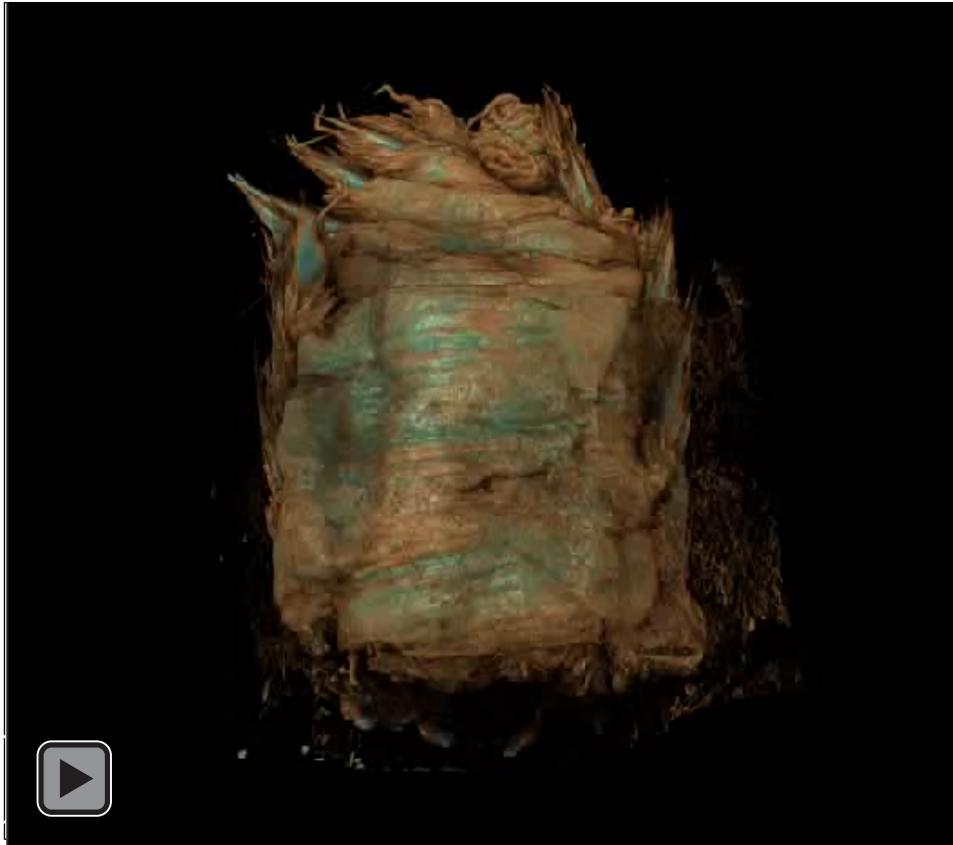


Figure 15. *Hermodice carunculata*, false-colour volume rendering of **a** anterior region, antero-dorsal view **b** virtually dissected anterior end showing pharyngeal structures, lateral view **c** virtually dissected anterior end, lateral view **d** virtually dissected anterior end at level of mouth opening, anterior view.

and X-rays were absorbed by the tissues according to their natural density differences and atomic number. Mineralised structures (jaw apparatus, chaetae) showed up most clearly, as well as the vascular system in the cephalic part of the animal. Muscle tissues are less dense but clearly visible, whereas nervous tissues (brain, ganglia) have almost no X-ray attenuation. Through careful observation, almost all anatomical features can be observed in the data (Video 1); however, external features are partly obscured by a thin ethanol film clinging to the specimen during scanning. By drying the specimen with HMDS, such artefacts can be avoided, since the specimen is fully desiccated while retaining its morphology. (e.g. Figs 7, 15). The process removes the liquid medium both from the cells and the surrounding area, resulting in a sharp overall contrast of all tissues. However, density variations between different tissues are less pronounced than in the wet specimen, creating difficulties in distinguishing neighbouring organs (Figs 10a, 15c–d). Scanning specimens in air can create other artefacts caused by slight vibration of protruding structures during rotation, such as the chaetae in the parapodium of *Eunice roussaei* (Fig. 7d). These artefacts become more pronounced with increasing magnification, since the effect of the movement becomes stronger.

Iodine and PTA bind to tissues and thus increase their X-ray energy absorption rate. The two stains generally bind to all tissues but exhibited different affinities to certain tissue types, staining them more intensely. Iodine seems to stain calcified struc-



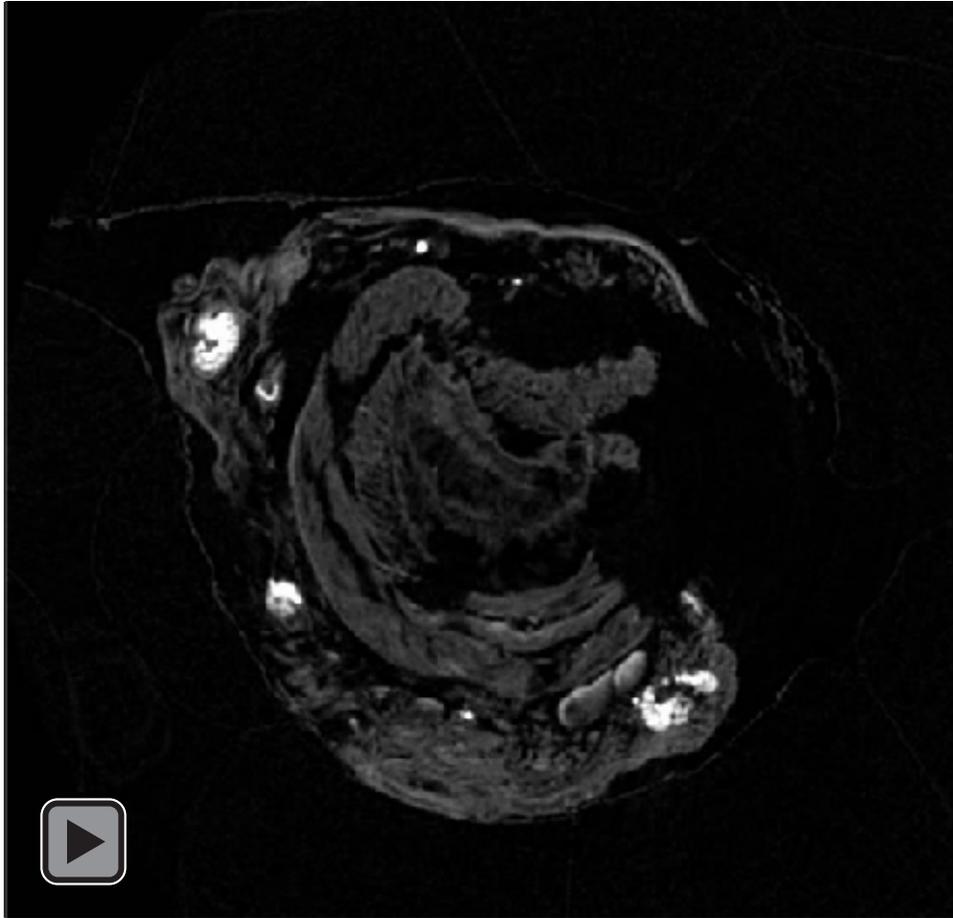
Video 3. *Hermodice carunculata*, false-colour volume rendering, virtual dissection of anterior end, dorsal view. Video available for download in full resolution from <http://polychaetes.marbigen.org/hermodice-carunculata-micro-ct-video>.

tures and polysaccharides more strongly, whereas phosphotungstic acid is known to bind to certain proteins (fibrin, collagen) (Quintarelli et al. 1973) and stains the cuticle and muscle tissues more intensely. In the *Eunice* specimen stained with iodine, the cuticle (containing polysaccharides), the jaw apparatus (mineralised with aragonite) and the subacicular hooks were stained strongly compared to other tissues (Fig. 7b). The chemical composition of the subacicular hooks is unknown, but calcified chaetae are known to occur in other polychaetes (e.g. Amphinomidae (Westheide 1997), Pogonophora (George and Southward 1973)), thus their strong staining with iodine could indeed be caused by calcium components. In *Syllis gracilis*, the proventricle showed an increased density after staining, here the muscle fibres likewise contain calcium (Briggs et al. 1985) (Fig. 14). In the *Eunice* specimen stained with PTA, the cuticle, which consists of collagen fibres, was stained very intensely, especially in the anterior region of the animal (Fig. 7a). Muscular fibres surrounding the jaw apparatus, the longitudinal muscles and the parapodial muscles are likewise more pronounced with PTA than

with iodine (not shown). Contrary to the iodine-stained specimen, the jaw apparatus and the subacicular hooks of the PTA-stained specimen were not stained and remain almost invisible (Fig. 7a). In the two phyllodocid specimens, PTA worked exceptionally well on the pharyngeal apparatus (Figs 10–11). In larger specimens, such as in *Alitta succinea*, the stain does not easily penetrate into the tissue—black areas in the images indicate unstained tissue (Fig. 9f). PTA penetrates slowly and tissue can bind large volumes of the chemical, so renewal every few days (or larger amounts) of the solution and longer staining times are required for larger specimens.

Visualisation of results

Different approaches to communicate three-dimensional data through a scientific publication have been explored in this study. Firstly, two-dimensional images (screenshots) have been created from the volume rendering software, both normal images and stereo view images which, when viewed with red-cyan glasses, create a 3D-effect. Secondly, videos of interaction with the three-dimensional data have been created and directly embedded into the PDF version of this article. Thirdly, three-dimensional models have been embedded into the PDF which, when viewed with an Adobe Acrobat product, allow the user to interact with them (e.g. rotating, zooming). Both images and videos communicate predefined views of the data. Videos, however, contain a substantially larger amount of information than a single static image. The video of the sequence of cross sections through the data (Video 4) allows the user to investigate the full dataset in a very compact version, thus information is conveyed which would be impossible to include in a publication if only images were used. In other cases, videos allow the viewer to better perceive the spatial relation of structures to each other and understand their relative position and perhaps their functioning in the organism (Videos 1–3). The interactive models give the reader the greatest freedom to explore the data. Surface models provide an excellent method to present selected information on specific structures (e.g. jaw apparatus, parapodium with chaetae (Figs 2, 8)). The resulting models can be rotated, magnified and individual parts can be isolated, allowing the user to explore shapes and spatial relationships from all angles. However, surface models are more suitable for compact structures such as internal organs (see e.g. Ruthensteiner and Hess 2008) than for the display of fine details. In models with complex shapes and details the number of vertices increased dramatically, leading to very large file sizes which were not suitable for further processing. The relatively complex shape of the parapodium with branchiae is already at the limit of what a current standard desktop computer can process. Smoothing the surface and reduction of vertices reduced the data significantly, but automated smoothing or reduction risks eliminating small but taxonomically important details. The limbate and pectinate chaetae had to be omitted from the model since their thin structures were reduced to random dots along their length as soon as the number of vertices was reduced. The compound chaetae had to be completely remodelled and were finally included as simplified shapes, showing general



Video 4. *Hermodice carunculata*, sequence of transaxial cross sections from post-pharyngeal chaetigers to the prostomium. Video available for download in full resolution from <http://polychaetes.marbigen.org/hermodice-carunculata-slices-micro-ct-video>.

shape and position but no details of structure. Manual post-processing of the model produced better accounts but requires detailed original data and a good knowledge of the morphology of the specimen. The process can furthermore become very time-consuming and the effort/outcome ratio has to be carefully considered. In the data produced during this study a hybrid approach was used: for most structures manual processing was considered not necessary, some characters were re-modelled as simplified shapes, others omitted. The optimal balance between model size and conveyed information will always depend on the research question being addressed. Software approaches, such as SPIERS (Sutton et al. 2012) that optimise the rendering of models with a high number of vertices can be an excellent solution if simplified shapes should be presented to the user.

To include the interactive volume data, the dataset had to be substantially reduced in size, resulting in a loss of many details. The embedded data are not a true volume rendering (based on a stack of images) but a pseudo-volume rendering: the software exports images from each angle of the volume-rendered objects and presents these to the user, creating the illusion of a three-dimensional object. The transfer function (and thus the information content) are predefined during model creation and cannot be changed by the end-user. The resulting model has thus not only limitations concerning the available detail of data but also towards the options for the user to explore the data, rotating and zooming being the only options of interactivity. The available transfer functions are furthermore not as sophisticated as in a desktop software. The slight density differences between the muscle tissue of the jaw apparatus and certain structures of the maxillary apparatus (maxillary carriers, accessory lamellae, MV) could not be visualised with the S2PLOT library (Fig. 5). The information value of these embedded objects is therefore limited, and in the present data the surface model is actually able to convey more information than the volume rendering.

Molecular analyses

No differences between the 16s rRNA sequences from samples before and after scanning could be detected; moreover, the 16s rRNA sequences of samples with increasing exposure time to X-ray radiation were also identical. GenBank accession numbers of sequences before scanning and after different radiation energy and exposure time are listed in Table 2.

Table 2. GenBank accession numbers of the sequences obtained from *Hediste diversicolor* specimens before and after scanning.

Sample code	Scanning time and voltage	GenBank accession number
NER015	none (control)	KC113440
NER015	1.5 h, 100 kV	KC113442
NER063	none (control)	KC113441
NER063	12 h, 60 kV	KC113443
NER063	24 h, 60 kV	KC113444
NER063	36 h, 60 kV	KC113445

Discussion

Is micro-computed tomography suitable for the creation of cybertypes?

Accuracy and reliability of information:

Three-dimensional data resulting from micro-CT contain a wealth of information for systematists and taxonomists. The examination of characters in their natural position within the organism allows researchers to assess their true shape but also to infer func-

tionality from morphological structures or even discover new diagnostic characters (e.g. Ziegler et al. 2010, Zimmermann et al. 2011). The non-destructive character of the technology allows examining internal features (even of very dense materials) without the need for dissection, leaving the original material intact. At the same time, characters can be investigated truly in three dimensions, whereas dissected material often has to be manipulated or squeezed for its adequate observation under a light microscope, thus rearranging the position of characters. The digital data allow the accurate identification of characters in three dimensions and relate them to each other. If needed, parts of the data can be isolated and examined separately, allowing their free manipulation without other body parts obscuring them (e.g. Figs 4b–c). With appropriate software, measurements can be performed in three dimensions, paving the way for a greater accuracy in these analyses which cannot be obtained with conventional methods.

Despite these obvious advantages, there are also certain limitations of micro-CT. A crucial point for research on small-sized organisms is image resolution which currently lies—depending on the system—in the range of 0.8 to 100 $\mu\text{m}/\text{pixel}$ (although nano-CT systems can reach 0.1 $\mu\text{m}/\text{pixel}$ (Wang et al. 2008)). This is a far coarser resolution than other techniques such as scanning electron microscopy, histological sectioning, cLSM or even light microscopy can achieve. Depending on the size of the specimen and the taxonomically important characters of the taxon, this resolution can prove sufficient. In many taxa, however, fine details in the micrometre range are important characters to distinguish species. These structures might be at the resolution limit of current micro-CT desktop scanners and cannot be adequately displayed. In some cases, the proper observation of characters might also be restricted due to their positioning during the scan. Under a stereo microscope specimens can be twisted or stretched, and obscured features can be made visible by careful manipulation with forceps. The data resulting from X-ray scanning are of course static, since only one position of the animal is imaged in a scan. Virtual dissections or measurements can partly overcome the problems, but in some cases the desired feature might simply not be visible in the data. Another limitation of micro-computed tomography is its inability to detect true colours. Pigmentation patterns can contain valuable taxonomic information in many taxa, but with current systems this information is not available. Recent developments of hybrid systems combining micro-CT with photon counting (Qiong et al. 2012) show, however, the first promising results of true-colour micro-CT imaging and might become a standard component of desktop scanners in the future. Until such hybrid systems become widely available, a solution for creating true-colour 3D models would be to use photographs or laser scans of the organisms and wrap these around the surface models to create realistic surfaces. However, these options require a substantial amount of manual post-processing and familiarity with 3D creation software.

Apart from inherent limitations of the technique, the information value of a dataset also depends on a range of parameters and settings during the image production process. Besides artefacts which might be created during image acquisition (Abel et al. 2012), each step from the acquisition to the final presentation presents the user with a number of settings and choices, each of which can influence the final results (Fig. 16) and cre-

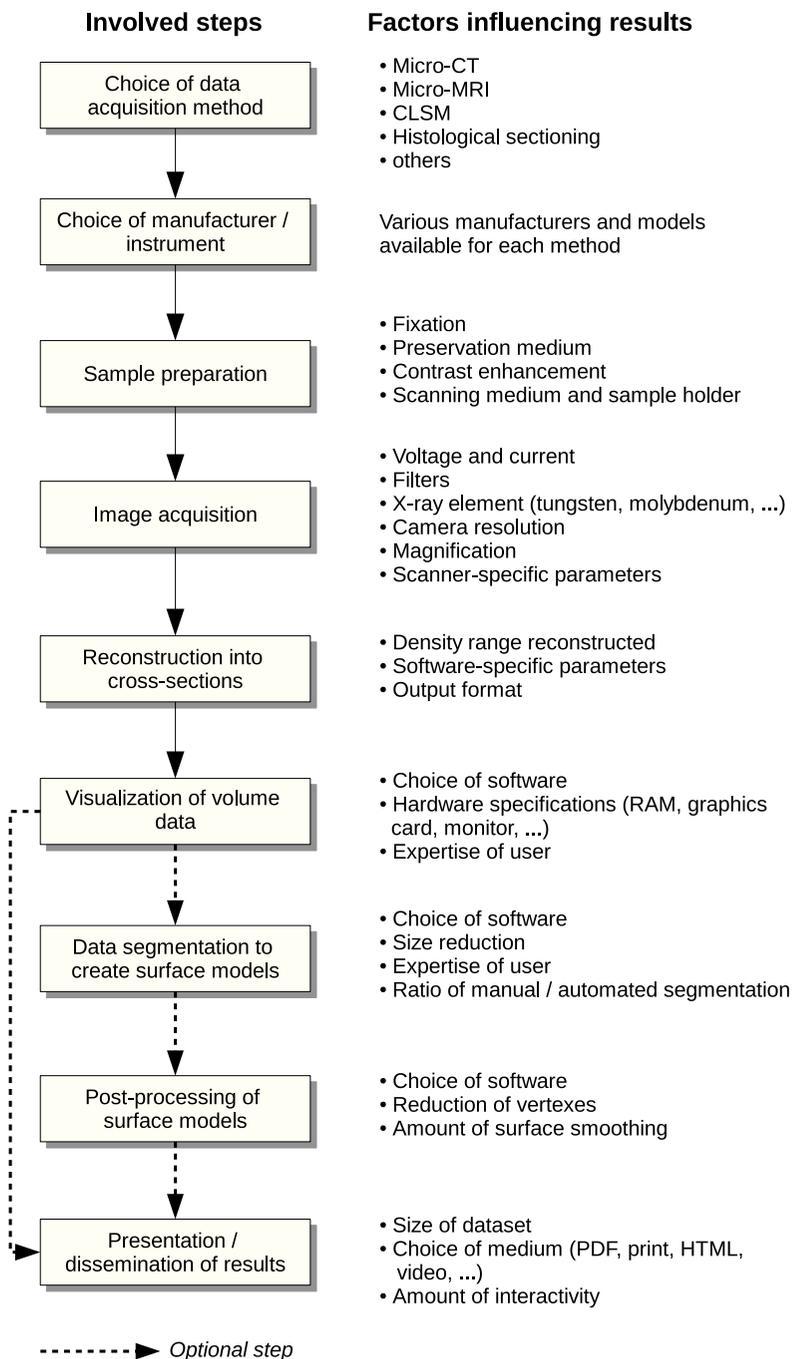


Figure 16. Diagram of the image acquisition process from the choice of method to the final presentation of the data, including factors influencing the outcome and information value of the results.

ate problems of comparability between datasets. Since at each step a level of subjectivity is added, the closest approximate to raw data in this chain—the initial volumetric data—are most suitable to represent a virtual type, although the settings during sample preparation, scanning and reconstructing already determine much of the information content. By this definition, derivative data products, such as the interactive surface and volumetric models included in this publication would not qualify for a cybertype. However, the ultimate target which governs these settings and choices is to obtain the optimum result and an adequate view of the character(s) in question. In disciplines other than zoology, different conventions might be more adequate (e.g. Sutton et al. 2012).

Effects on physical specimens:

Micro-CT is commonly characterised as a non-destructive imaging technique, and indeed neither does the specimen need to be physically manipulated before scanning nor does the exposure to X-ray energies have any visible effect on the morphology of the specimen. However, nothing is known yet about possible tissue damage at the cellular level or after elongated or repeated exposure, so whether the technique is absolutely non-destructive remains to be proved. A certain risk of altering the specimen's characters lies in their preparation for image acquisition, specifically in methods for contrast enhancement which might irreversibly change tissue characteristics. Although micro-CT does not *per se* require contrast enhancement, such treatment might be necessary, especially when scanning soft-bodied organisms. Tissue staining can lead to excellent results and sharp image contrast (Metscher 2009a, b) but no universally applicable protocols for the removal of these stains exist so far. The selectivity of stains towards different tissue types particularly enhances the contrast of certain tissues, whereas signals of other tissues are suppressed and show up less clearly or not at all in the scans. This alteration of natural tissue contrasts renders the specimen potentially useless for future micro-CT examinations with a different purpose. Tissue staining should therefore be used with care, especially if the data are intended for general purpose studies and/or if the material is valuable, since the long-term effects of these chemicals remain yet unknown. On the contrary, specimens scanned in air (either wet or dried) can simply be immersed again into the preservation medium after scanning and are thus available both for imaging methods and for investigation through traditional microscopy techniques, but again not all samples are suitable for this type of treatment. The absolute impact of each method is almost impossible to predict due to the large diversity of tissue types, chemical components and material combinations existing in invertebrates (Faulwetter et al. 2012). Protocols for best practice have, therefore, still to be identified experimentally for each species. The most suitable method is always determined by a number of factors such as the characteristics of the sample (density, size, shape), the surrounding medium and the scope of the study (Table 3). As far as indicated by the results of the current study, X-ray radiation induced by micro-CT seems not to have affected the molecular identity of the specimens, at least no effects of the radiation on the sequenced fragment of the 16S rRNA gene could be detected. Neither exposure to high energies nor repetitive

Table 3. Overview of different contrast enhancing techniques, their applications and limitations

Method	Use cases	Limitations	Reversibility
Removal of preservation medium, scan in humid state	<ul style="list-style-type: none"> Natural contrast between different tissue types should be kept Specimen cannot be stained Specimen contains dense parts and only these should be visualised 	<ul style="list-style-type: none"> Remaining drops of liquid between external features might obscure details Specimens might dry out if scanning time is too long External body parts such as branchiae, membranes, might collapse into unrecognisable shapes Soft tissue inside very dense structures cannot be visualised appropriately 	Immerse into preservation medium
Drying (HMDS, possibly also critical point drying or freeze drying)	<ul style="list-style-type: none"> Specimen contains both dense parts and soft tissue and both should be visualised Specimen cannot be stained nor scanned in a humid state 	<ul style="list-style-type: none"> Specimens become fragile, external body parts might break off Long appendages vibrate during rotation and create blur Natural contrast between tissues is reduced, individual organs might be difficult to separate 	Immerse in >95% ethanol
Tissue stains (e.g. iodine, PTA, silver staining)	<ul style="list-style-type: none"> Specimen contains both dense parts and soft tissue and both should be visualised Specimen is very small or fragile Specimen cannot be removed from liquid medium Specimen has appendages that could vibrate in air or collapse to body when liquid medium is removed Only certain tissues should be visualised (selective staining) 	<ul style="list-style-type: none"> Natural contrast between tissues is reduced, individual organs might be difficult to separate Tissues stain selectively, some tissues might not show up at all in the image PTA: large specimens need very long staining times and large amounts of stain Iodine: soft tissue does not stain well when specimen contains large calcareous structures (e.g. mollusc shells, serpulid tubes) 	unknown

exposure seems to have caused any mutations in this specific part of the molecular material. Previous studies attempting to determine whether exposure of preserved tissue to X-ray radiation causes a fragmentation of the DNA have reached contradicting conclusions (Götherstrom et al. 1995, Grieshaber et al. 2008, Paredes et al. 2012). As Paredes et al. (2012) point out, the findings of Götherstrom et al. (1995) and Grieshaber et al. (2008) might however be biased by the setup of the experiment. Götherstrom et al. (1995) examined the effect of X-ray radiation on the ability to amplify DNA fragments from pig bones, basing their conclusions on the results on the differences of PCR products brightness in an agarose gel. Results of Grieshaber et al. (2008) did not show any significant effect of the radiation on the DNA amplification with RT-PCR; nevertheless both studies conclude that exposure to radiation had caused degradation of the DNA. Paredes et al. (2012), by comparing pre- and post-CT DNA fragmentation profiles of preserved bird skin, did not detect significant differences of DNA quality before and after scanning, concluding that no quantifiable DNA fragmentation was induced by exposure of the sample to X-ray. However, all

studies use different exposure times and energies as well as different methods to detect the effect of radiation on the genetic material, thus all these results are only indicative. Furthermore, none of the previous studies used sequencing data to compare the X-Ray effect on DNA samples. If the methodology is to be applied more broadly, further experiments are clearly needed. Likewise, the possible effect of staining substances on the molecular material should be tested; however, DNA of specimens subjected to different contrast-enhancement methods could be successfully extracted and amplified in previous experiments (Austin and Dillon 1997, Faulwetter et al. 2012). The above aspects of the potential impact of the micro-CT technology on preserved biological material must be entirely resolved before this method is suggested for wide use on the specimens deposited in museums. This holds especially true for type-material of nominal species which is unique by definition.

Data access and curation of cybertypes

Three-dimensional data can be communicated in various ways and through various media. Their true potential lies, however, in the wealth of information that a full volumetric dataset offers to the skilled researcher. Volumetric datasets can easily reach a size of several gigabytes per dataset (Table 1), a fact that poses new challenges concerning the management, archival, backup and dissemination of these data. Currently, the community lacks infrastructures, standards and policies that allow the adequate curation of three-dimensional data (Rowe and Frank 2011). An urgent priority is the creation of an infrastructure for sharing these data and encouraging their reuse. Although a number of morphological databases exist at present (Ziegler et al. 2010), only few focus on the curation and visualisation of three-dimensional morphology data (e.g. Digimorph – <http://www.digimorph.org>), Digital Fish Library – <http://www.digitalfishlibrary.org>, Berquist et al. 2012), and none acts as a broad-scale repository of high-resolution volumetric datasets. Apart from these archives, hundreds of three-dimensional datasets of biological specimens have already been produced but remain inaccessible to the research community (Ziegler et al. 2011a, Boistel et al. 2011). The development of standards and protocols for archiving and disseminating three-dimensional data as well as the creation of centralised registers to make the information retrievable remains an immediate priority (Ziegler et al. 2010, Rowe and Frank 2011) and is crucial for the future success of these developments. In this context, natural history museums and other large natural collections will have to play a central role, not only by digitising their collections and thus massively producing data (as already exemplified by recent efforts (Smith and Blagoderov 2012)), but also by taking the lead in the development of standards and software for virtual museum collections and curation of cybertypes (for a discussion on the lack of standard file formats for interchange see Sutton et al. 2012). This leads to another important issue—the current lack of standards to properly document and exchange volumetric data. Without metadata, datasets are neither retrievable nor interpretable. The medical community has developed the

DICOM standard (Digital Imaging and Communications in Medicine, <http://dicom.nema.org/>) which contains both format definitions and a communication protocol for the description and exchange of volume data. However, this standard is not as universally implemented as one might expect, with different formats and versions being used by different parties (Mildenberger and Jensch 1999, Pianykh 2012). Furthermore, to what extent the DICOM specifications are in compliance with the purposes of the taxonomic community remains to be investigated, thus DICOM might or might not prove to be a suitable format for volumetric data definition and exchange. Another limiting factor for the optimal use of volumetric data is that powerful computing capabilities are required for the visualisation, exploration and analysis of large data sets. Although computing power is becoming increasingly more inexpensive, large-scale analyses and comparisons of datasets will likely be limited to high-performance computing centres. Virtual laboratories providing remote access to these facilities could however give a true boost to widespread usage of these data (Ziegler et al. 2010). System architectures to provide rapid access to three-dimensional data are already being developed by the informatics research community (e.g. Engel et al. 1999, Kaupp et al. 2002, Prohaska et al. 2004, Congote et al. 2009, 2012). Finally, a third challenge for the biomedical informatics community is the incorporation of three-dimensional information on web sites. A major bottleneck for the publication of volumetric data is still their large size which has to be transferred to the client for visualisation, but also the lack of native integration of this data into HTML standards. Developments such as the VAXML standard (Sutton et al. 2012) could potentially become the basis for such integrations. The near future will see several promising developments such as the Arivis WebView browser (<http://webview3d.arivis.com/>) or Voluminous, the web-based version of Drishti (under development, presentation at <https://sites.google.com/site/ozvizworkshop/ozviz-2011>).

The future of virtual taxonomy

The increasing availability of accurate, three-dimensional virtual representations of biological specimens offers an exciting range of new research opportunities and will significantly accelerate access to first-hand morphological information, thus helping to overcome one of the major bottlenecks in systematic and taxonomic research: the continuous availability of type material to all potential users simultaneously. At present, although virtual specimens in most cases cannot—and should not—replace physical type material, often the desired information can be obtained from a virtual representation. In this study, none of the scanned specimens comply to the hypothetical requirement for a virtual type to provide as much information as the original material. However, different imaging methods (e.g. photography, nano-CT, MRI, OPT) can be employed to produce complementary datasets, and sophisticated future methods could provide a seamless integration of different datasets, incrementally loading additional data when zooming in or focusing on certain characteristics. The availability

of information-rich cybertypes would not only protect the actual type material from loss or damage through careless handling, but would also provide simultaneous access by multiple users to the material. This approach would also provide a way to access collections where local restrictions prevent removing specimens from the institution or country.

The increased creation of three-dimensional taxonomic data will also inevitably influence the way taxonomic data is published. Embedding data as interactive, three-dimensional objects into publications will likely become a standard to usefully convey information. With the recent amendment of the International Code of Zoological Nomenclature (ICZN) that allows taxonomic treatments to be published exclusively electronically (International Commission on Zoological Nomenclature 2012), multimedia and interactive data can be embedded directly into the publication. This allows these large datasets to be embedded within the presentation of the paper as a single “coherent scientific report” (Ziegler et al. 2011b). With these regulations, an important milestone has been reached in the process of transforming taxonomy into a cyber-discipline. However, if the concept of virtual type material types is to be officially established, further changes to the ICZN will have to be made that regulate the use of cybertypes to complement physical type material. Specifically, Article 72, which regulates the type concept in nomenclature, should be amended to include the definition of the concept of a “cybertype”, “e-type” or “virtual type” and provide recommendations upon the nomenclatural status, electronic format, access and longevity of such datasets, as well as ensuring that they are made accessible to the public domain under a licence that ensures open access and encourages the creation of derivative works. Furthermore, as Wheeler et al. (2012) point out, regulations concerning registry of cybertypes in central access points such as ZooBank should be included. Such regulations will allow entirely new ways of data access to emerge, such as virtual realities (Deligiannidis and Jacob 2005, Laha et al. 2012), interactive access to virtual specimens on mobile devices (Johnson et al. 2012) or holographic representations (Javidi and Tajahuerce 2000, Midgley and Dunin-Borkowski 2009, Boistel et al. 2011).

Despite certain obvious advantages of 3D-imaging technologies, they will need time to evolve into a widely adopted method. At its core, taxonomy is a very traditional discipline, and commonly, changes are adopted at a slow pace. The aforementioned transition of taxonomy into a more data-centric and electronic discipline will need time so the community can learn how to make the best use of this new type of data and the information it contains, as well as to develop the necessary skills to handle these data. This goes hand in hand with technical obstacles that prevent the method from becoming widely used at the present: access to 3D imaging facilities is, although steadily increasing, still limited and often expensive. Special technical skills are needed to produce and process the data, and even with such expertise, the creation of the final dataset (the cybertype) is still a very time-consuming process. However, these arguments hold true for many new technologies, and it will be for the community to decide whether the information value contained in 3D-datasets will allow the technology to survive and to shape the future direction of taxonomy.

Conclusions

Up to now, morphology-based systematics and taxonomy have not been able to keep pace with the rapid developments and data creation that characterise other disciplines. New technologies, such as micro-CT and other imaging techniques, will allow massive, computer-accessible data production. This will, in turn, inspire the development of new tools to manage and analyse these data, allowing large-scale morphology-based phylogenies, semi-automated identifications, the formulation of new systematic hypotheses, and advanced research on novel ways of managing, visualising and publishing data. The combined efforts of humans and new technologies will help the discipline to find its way into the digital age and might trigger its renaissance with an impact rivaling the discoveries of the great naturalist era of the 19th century.

Acknowledgements

The authors are grateful for support from the following colleagues: Dr Costas Dounas (Hellenic Centre for Marine Research, Greece), Dr Didier Aurelle (Université de la Méditerranée Centre d'Océanologie de Marseille, France) and Dr Chariton Chintiroglou (Aristotelian University of Thessaloniki, Greece) for providing specimens; Dr Jeroen Hostens (SkyScan, Belgium) and Dr Brian Metscher (Department of Theoretical Biology, University of Vienna, Austria) for support and training in the application of various staining methods and scanning techniques; Dr Lyubomir Penev (PENSOFT, Sofia, Bulgaria) and Dr Gordon Paterson (Natural History Museum of London, UK) for inspiring discussions; Dr Farah Ahmed and Daniel Sykes (Natural History Museum of London, UK) for sharing experiences and for an introduction to Drishti, and finally Dr Antonios Magoulas (Hellenic Centre for Marine Research, Greece) for his continuous support through the MARBIGEN project. The reviewers (Dr Daniel Mietchen, Dr Russel Garwood and one anonymous reviewer) are thanked for their suggestions to improve the manuscript. Financial support was provided by the European Union projects MARBIGEN (FP7-REGPOT-2010-1) and ViBRANT (RI-261532).

References

- Abel RL, Laurini CR, Richter M (2012) A palaeobiologist's guide to "virtual" micro-CT preparation. *Palaeontologia Electronica* 15: 1–16. <http://palaeo-electronica.org/content/pdfs/284.pdf>
- Alba-Tercedor J, Sánchez-Tocino L (2011) The use of the SkyScan 1172 high-resolution micro-CT to elucidate if the spicules of the "sea slugs" (Mollusca: Nudibranchia, Opisthobranchia) have a structural or a defensive function. In: *MicroCT User Meeting Abstract Book*, Leuven (Belgium), April 2011. SkyScan, Kontich, Belgium, 113–121. http://www.skyscan.be/company/UM2011/abstract_22.pdf

- Audzijonyte A, Ovcarenko I, Bastrop R, Väinölä R (2008) Two cryptic species of the *Hediste diversicolor* group (Polychaeta, Nereididae) in the Baltic Sea, with mitochondrial signatures of different population histories. *Marine Biology* 155: 599–612. doi: 10.1007/s00227-008-1055-3
- Audouin JV, Milne-Edwards H (1834) Recherches pour servir a l'histoire naturelle du littoral de la France, ou recueil de mémoires sur l'anatomie, la physiologie, la classification et les moeurs des animaux de nos côtes, 2. Annélides, Crochard, Paris, 290 pp. <http://www.biodiversitylibrary.org/ia/recherchepourse21834audio>
- Austin AD, Dillon N (1997) Extraction and PCR of DNA from parasitoid wasps that have been chemically dried. *Australian Journal of Entomology* 36: 241–244. doi: 10.1111/j.1440-6055.1997.tb01461.x
- Bakken T (2004) A revision of Nereidinae (Polychaeta, Nereididae). PhD Thesis, Trondheim, Norway: Norwegian University of Science and Technology. <http://ntnu.diva-portal.org/smash/record.jsf?pid=diva2:123502>
- Barnes D, Fluke C, Bourke P, Parry OT (2006) An advanced, three-dimensional plotting library for astronomy. *Publications of the Astronomical Society of Australia* 23: 82–93. doi: 10.1071/AS06009
- Barroso R, Klautau M, Solé-Cava AM, de Paiva PC (2009) *Eurythoe complanata* (Polychaeta: Amphinomidae), the “cosmopolitan” fireworm, consists of at least three cryptic species. *Marine Biology* 157: 69–80. doi: 10.1007/s00227-009-1296-9
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2005) GenBank. *Nucleic Acids Research* 33: D34–D38. doi: 10.1093/nar/gki063
- Berquist RM, Gledhill KM, Peterson MW, Doan AH, Baxter GT, Yopak KE, Kang N, Walker HJ, Hastings PA, Frank LR (2012) The Digital Fish Library: Using MRI to Digitize, Database, and Document the Morphological Diversity of Fish. *PLoS ONE* 7: e34499. doi: 10.1371/journal.pone.0034499
- Boistel R, Swoger J, Kržič U, Fernandez V, Gillet B, Reynaud EG (2011) The future of three-dimensional microscopic imaging in marine biology. *Marine Ecology* 32: 438–452. doi: 10.1111/j.1439-0485.2011.00442.x
- Bond J, Hörnschmeyer T, Young P, Yang L (2008) Meshing of XMT data to study the functional morphology of beetle *Priacma serrata*'s head and mandible. In: Proceedings of the 2nd International Conference of Bionic Engineering (ICBE'08), Changchun (China) October 2008. Jilin University, Changchun, 1–11.
- Briggs RT, Chaffee JE, Anderson M (1985) Calcium-containing granules in myoepithelial cells of the polychaete *Syllis spongiphila*: Possible ionic modulators. *Tissue* 17: 923–928. doi: 10.1016/0040-8166(85)90046-1
- Budd GE, Olsson L (2007) A renaissance for evolutionary morphology. *Acta Zoologica* 88: 1. doi: 10.1111/j.1463-6395.2007.00262.x
- Böttgermann M (2002) Revision of the Glyceridae Grube 1850 (Annelida: Polychaeta). *Abhandlungen der Senckenbergischen Naturforschenden Gesellschaft* 555: 1–249.
- Carrera-Parra LF (2006a) Revision of *Lumbrineris* de Blainville, 1828 (Polychaeta: Lumbrineridae). *Zootaxa* 1336: 1–64. <https://www.scopus.com/record/display.url?eid=2-s2.0-33750323441&origin=resultslist>

- Carrera-Parra LF (2006b) Phylogenetic analysis of Lumbrineridae Schmarda, 1861 (Annelida: Polychaeta). *Zootaxa* 36: 1–36. <https://www.scopus.com/record/display.url?eid=2-s2.0-33750375288&origin=resultslist>
- Carvalho MR, Bockmann FA, Amorim DS, Brandão CRF, Vivo M, Figueiredo JL, Britski HA, Pinna MCC, Menezes NA, Marques FPL, Papavero N, Cancellato EM, Crisci JV, McEachran JD, Schelly RC, Lundberg JG, Gill AC, Britz R, Wheeler QD, Stiassny MLJ, Parenti LR, Page LM, Wheeler WC, Faivovich J, Vari RP, Grande L, Humphries CJ, DeSalle R, Ebach MC, Nelson GJ (2007) Taxonomic Impediment or Impediment to Taxonomy? A Commentary on Systematics and the Cybertaxonomic-Automation Paradigm. *Evolutionary Biology* 34: 140–143. doi: 10.1007/s11692-007-9011-6
- Claparède É (1870) Les annélides chétopodes du Golfe de Naples. Supplément. *Mémoires de la Société de physique et d'histoire naturelle de Genève* 20: 365–542. <http://www.biodiversitylibrary.org/item/18576>
- Congote J, Kabongo L, Moreno A, Segura A, Beristain A, Posada J, Ruiz O (2012) Volume Ray Casting in WebGL. In: Mukai N (Ed) *Computer Graphics. InTech*, Rijeka, 157–178. doi: 10.5772/34878
- Congote J, Kabongo L, Moreno A, Segura A, Posada J, Ruiz O (2009) Interactive visualization of volumetric data with WebGL in real-time. In: *Web3D '11 Proceedings of the 16th International Conference on 3D Web Technology*, Paris (France), June 2011. ACM New York, 137–146. doi: 10.1145/2010425.2010449
- Cook LG, Edwards RD, Crisp MD, Hardy NB (2010) Need morphology always be required for new species descriptions? *Invertebrate Systematics* 24: 322–326. doi: 10.1071/IS10011
- Csösz S (2012) Nematode infection as significant source of unjustified taxonomic descriptions in ants (Hymenoptera: Formicidae). *Myrmecological News* 17: 27–31. http://www.myrmecologicalnews.org/cms/images/pdf/volume17/mn17_27-31_non-printable.pdf
- Deans AR, Yoder MJ, Balhoff JP (2011) Time to change how we describe biodiversity. *Trends in Ecology & Evolution* 27: 78–84. doi: 10.1016/j.tree.2011.11.007
- Deligiannidis L, Jacob RJK (2005) An Immersive Environment for the Vase Museum. In: *HCI'05 International Conference on Human-Computer Interaction*, Las Vegas (USA), July 2005. HCI International, 54–62. <https://www.scopus.com/record/display.url?eid=2-s2.0-60749101560&origin=resultslist>
- Dierick M, Cnudde V, Masschaele B, Vlassenbroeck J, Van Hoorebeke L, Jacobs P (2007) Micro-CT of fossils preserved in amber. *Nuclear Instruments and Methods in Physics Research A* 580: 641–643. doi: 10.1016/j.nima.2007.05.030
- Dinley J, Hawkins L, Paterson GLJ, Ball AD, Sinclair I, Sinnett-Jones P, Lanham S (2010) Micro-computed X-ray tomography: a new non-destructive method of assessing sectional, fly-through and 3D imaging of a soft-bodied marine worm. *Journal of Microscopy* 238: 123–33. doi: 10.1111/j.1365-2818.2009.03335.x
- Dunlop JA, Wirth S, Penney D, McNeil A, Bradley RS, Withers PJ, Preziosi RF (2011) A minute fossil phoretic mite recovered by phase-contrast X-ray computed tomography. *Biology Letters* 8: 457–460. doi: 10.1098/rsbl.2011.0923
- Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sørensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kris-

- tensen RM, Wheeler WC, Martindale MQ, Giribet G (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452: 745–749. doi: 10.1038/nature06614
- Edgecombe GD, Giribet G, Dunn CW, Hejzol A, Kristensen RM, Neves RC, Rouse GW, Worsaae K, Sørensen MV (2011) Higher-level metazoan relationships: recent progress and remaining questions. *Organisms Diversity & Evolution* 11: 151–172. doi: 10.1007/s13127-011-0044-4
- Engel K, Sommer O, Ernst C, Ertl T (1999) Remote 3D Visualization using Image-Streaming Techniques. In: Proceedings of the 11th International Conference on Systems Research, Informatics and Cybernetics, Baden-Baden (Germany), August 1999. International Institute for Advanced Studies in Systems Research and Cybernetics, 91–96.
- Evenhuis NL (2007) Helping Solve the “other” Taxonomic Impediment: Completing the Eight Steps to Total Enlightenment and Taxonomic Nirvana. *Zootaxa* 1407: 3–12. <https://www.scopus.com/record/display.url?eid=2-s2.0-33847357817&origin=resultslist>
- Faulwetter S, Dailianis T, Vasileiadou A, Arvanitidis C (2012) Investigation of contrast enhancing techniques for the application of Micro-CT in marine biodiversity studies. In: SkyScan MicroCT User meeting 2012. SkyScan, Brussels, Belgium, 12–20. <http://www.skyscan.be/company/UM2012/02.pdf>
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase unit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3: 294–299.
- George JD, Southward EC (1973) A comparative study of the setae of Pogonophora and polychaetous Annelida. *Journal of the Marine Biological Association of the United Kingdom* 53: 397–401. doi: 10.1017/S0025315400022359
- Giribet G (2010) A new dimension in combining data? The use of morphology and phylogenomic data in metazoan systematics. *Acta Zoologica* 91: 11–19. doi: 10.1111/j.1463-6395.2009.00420.x
- Godfray HCJ (2002a) Challenges for taxonomy. *Nature* 417: 17–19. doi: 10.1038/417017a
- Godfray HCJ (2002b) Towards taxonomy’s “glorious revolution”. *Nature* 420: 461. doi: 10.1038/420461a
- Godfray HCJ (2007) Linnaeus in the information age. *Nature* 446: 259–260. doi: 10.1038/446259a
- Golding RE, Jones AS (2006) Micro-CT as a novel technique for 3D reconstruction of molluscan anatomy. *Molluscan Research* 27: 123–128. <https://www.scopus.com/record/display.url?eid=2-s2.0-36048997602&origin=resultslist>
- Golding RE, Ponder WF, Byrne M (2009) Three-dimensional reconstruction of the odontophoral cartilages of Caenogastropoda (Mollusca: Gastropoda) using micro-CT: Morphology and phylogenetic significance. *Journal of Morphology* 270: 558–587. doi: 10.1002/jmor.10699
- Götherstrom A, Fischer C, Linden K (1995) X-raying ancient bone: a destructive method in connection with DNA analysis. *Laborativ Arkeologi* 8: 26–28.
- Grieshaber BM, Osborne DL, Doubleday AF, Kaestle FA (2008) A pilot study into the effects of X-ray and computed tomography exposure on the amplification of DNA from bone. *Journal of Archaeological Science* 35: 681–687. doi: 10.1016/j.jas.2007.06.001

- Grube AE (1840) Actinien, Echinodermen und Würmer des Adriatischen- und Mittelmeers, nach eigenen Sammlungen beschrieben. J.H. Bon, Königsberg, 92 pp. <http://www.biodiversitylibrary.org/bibliography/10133>
- Heim I, Nickel M (2010) Description and molecular phylogeny of *Tethya leysae* sp. nov. (Porifera, Demospongiae, Hadromerida) from the Canadian Northeast Pacific with remarks on the use of microtomography in sponge taxonomy. *Zootaxa* 21: 1–21. <https://www.scopus.com/record/display.url?eid=2-s2.0-77953923313&origin=resultslist>
- Hendrickx H, Cnudde V, Masschaele B, Dierick M, Vlassenbroeck J, Van Hoorebeke L (2006) Description of a new fossil Pseudogarypus (Pseudoscorpiones: Pseudogarypidae) with the use of X-ray micro-CT to penetrate opaque amber. *Zootaxa* 1305: 41–50. <https://www.scopus.com/record/display.url?eid=2-s2.0-33748599256&origin=resultslist>
- Holford P (2008) X-ray computerised microtomography (MicroCT): a new technique for assessing external and internal morphology of bees. *Journal of Apicultural Research* 47: 286–291. doi: 10.3896/IBRA.1.47.4.09
- Huckstorf K, Wirkner CS (2011) Comparative morphology of the hemolymph vascular system in krill (Euphausiacea; Crustacea). *Arthropod structure & development* 40: 39–53. doi: 10.1016/j.asd.2010.07.004
- International Commission on Zoological Nomenclature (2012) Amendment of Articles 8, 9, 10, 21 and 78 of the International Code of Zoological Nomenclature to expand and refine methods of publication. *ZooKeys* 219: 1–10. doi: 10.3897/zookeys.219.3944
- Javidi B, Tajahuerce E (2000) Three-dimensional object recognition by use of digital holography. *Optics Letters* 25: 610–612. doi: 10.1364/OL.25.000610
- Johnson PT, Zimmerman SL, Heath D, Eng J, Horton KM, Scott WW, Fishman EK (2012) The iPad as a mobile device for CT display and interpretation: diagnostic accuracy for identification of pulmonary embolism. *Emergency Radiology* 19: 323–327. doi: 10.1007/s10140-012-1037-0
- Joppa LN, Roberts DL, Pimm SL (2011a) The population ecology and social behaviour of taxonomists. *Trends in Ecology & Evolution* 26: 551–553. doi: 10.1016/j.tree.2011.07.010
- Joppa LN, Roberts DL, Pimm SL (2011b) Taxonomy that matters: Response to Bacher. *Trends in Ecology & Evolution* 27: 66. doi: 10.1016/j.tree.2011.11.003
- Kaup G, Stepanenko S, Herrmann A (2002) World-wide interaction with 3D-data. In: SS-GRR2002S, L'Aquila (Italy), July/August 2002. SSGRR, 1–11.
- Kersemans V, Thompson J, Cornelissen B, Woodcock M, Allen PD, Buls N, Muschel RJ, Hill MA, Smart SC (2011) Micro-CT for anatomic referencing in PET and SPECT: radiation dose, biologic damage, and image quality. *The Journal of Nuclear Medicine* 52: 1827–1833. doi: 10.2967/jnumed.111.089151
- Knapp S, Bateman RM, Chalmers NR, Humphries CJ, Rainbow PS, Smith AB, Taylor PD, Vane-Wright RI, Wilkinson M (2002) Taxonomy needs evolution, not revolution. *Nature* 419: 559. doi: 10.1038/419559a
- Kumar P, Ziegler A, Grahn A, Hee CS, Ziegler A (2010) Leaving the structural ivory tower, assisted by interactive 3D PDF. *Trends in Biochemical Sciences* 35: 419–422. doi: 10.1016/j.tibs.2010.03.008

- Laforsch C, Imhof H, Sigl R, Settles M, Heß M, Wanninger A (2012) Applications of Computational 3D–Modeling in Organismal Biology. In: Alexandru C (Ed) Modeling and Simulation in Engineering. InTech, Rijeka, 117–142. doi: 10.5772/31092
- Laha B, Sensharma K, Schiffbauer JD, Bowman DA (2012) Effects of immersion on visual analysis of volume data. IEEE transactions on visualization and computer graphics 18: 597–606. doi: 10.1109/TVCG.2012.42
- Leuckart R (1847) Verzeichniss der zur Fauna Helgolands gehörenden wirbellosen Seethiere. In: Frey and Leuckart, Beiträge zur Kenntniss wirbelloser Thiere mit besonderer Berücksichtigung der Fauna des Norddeutschen Meeres. Vieweg, Braunschweig, 136–168. <http://www.biodiversitylibrary.org/bibliography/2128>
- Marxen JC, Prymak O, Beckmann F, Neues F, Epple M (2007) Embryonic shell formation in the snail *Biomphalaria glabrata*: a comparison between scanning electron microscopy (SEM) and synchrotron radiation micro computer tomography (SR CT). Journal of Molluscan Studies 74: 19–26. doi: 10.1093/mollus/eym044
- McPeck MA, Symes LB, Zong DM, McPeck CL (2011) Species recognition and patterns of population variation in the reproductive structures of a damselfly genus. Evolution 65: 419–428. doi: 10.1111/j.1558-5646.2010.01138.x
- Metscher BD (2009a) MicroCT for comparative morphology: simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues. BMC Physiology 9: 11. doi: 10.1186/1472-6793-9-11
- Metscher BD (2009b) MicroCT for developmental biology: a versatile tool for high-contrast 3D imaging at histological resolutions. Developmental Dynamics 238: 632–640. doi: 10.1002/dvdy.21857
- Midgley PA, Dunin-Borkowski RE (2009) Electron tomography and holography in materials science. Nature Materials 8: 271–280. doi: 10.1038/nmat2406
- Mildenberger P, Jensch P (1999) Implementation of the DICOM standard in a heterogeneous environment: incompatibility or interoperability? Der Radiologe 39: 282–285. doi: 10.1007/s001170050510
- Molineux A, Scott RW, Ketcham RA, Maisano JA (2007) Rudist taxonomy using X-ray computed tomography. Palaeontologica Electronica 10: 13. http://palaeo-electronica.org/2007_3/135/index.html
- Nickel M, Donath T, Schweikert M, Beckmann F (2006) Functional morphology of *Tethya* species (Porifera): 1. Quantitative 3D-analysis of *Tethya wilhelma* by synchrotron radiation based X-ray microtomography. Zoomorphology 125: 209–223. doi: 10.1007/s00435-006-0021-1
- Nygren A, Eklöf J, Pleijel F (2010) Cryptic species of *Notophyllum* (Polychaeta: Phyllodocidae) in Scandinavian waters. Organisms Diversity & Evolution 10: 193–204. doi: 10.1007/s13127-010-0014-2
- Nygren A, Pleijel F (2011) From one to ten in a single stroke – resolving the European *Eumida sanguinea* (Phyllodocidae, Annelida) species complex. Molecular phylogenetics and evolution 58: 132–141. doi: 10.1016/j.ympev.2010.10.010
- Osborn KJ, Rouse GW (2010) Phylogenetics of Acrocirridae and Flabelligeridae (Cirratuliformia, Annelida). Zoologica Scripta 40: 204–219. doi: 10.1111/j.1463-6409.2010.00460.x

- Pallas PS (1766) *Miscellanea zoologica. Quibus novae imprimis atque obscurae animalium species describuntur et observationibus iconibusque illustrantur.* Petrum van Cleef, Hagí Comitum, 244 pp. <http://books.google.gr/books?id=cTQVAAAAQAAJ>
- Paredes UM, Prys-Jones R, Adams M, Groombridge J, Kundu S, Agapow P-M, Abel RL (2012) Micro-CT X-rays do not fragment DNA in preserved bird skins. *Journal of Zoological Systematics and Evolutionary Research* 50: 247–250. doi: 10.1111/j.1439-0469.2012.00657.x
- Patek SN, Nowroozi BN, Baio JE, Caldwell RL, Summers AP (2007) Linkage mechanics and power amplification of the mantis shrimp's strike. *The Journal of Experimental Biology* 210: 3677–3688. doi: 10.1242/jeb.006486
- Penev L, Agosti D, Georgiev T, Catapano T, Miller J, Blagoderov V, Roberts D, Smith VS, Brake I, Rycroft S, Scott B, Johnson NF, Morris R a., Sautter G, Chavan V, Robertson T, Remsen D, Stoev P, Parr C, Knapp S, Kress WJ, Thompson FC, Erwin T (2010) Semantic tagging of and semantic enhancements to systematics papers: ZooKeys working examples. *ZooKeys* 50: 1–16. doi: 10.3897/zookeys.50.538
- Penney D, Dierick M, Cnudde V, Masschaele B, Vlassenbroeck J, Van Hoorebeke, L, Jacobs P (2007) First fossil Micropholcommatidae (Araneae), imaged in Eocene Paris amber using X-Ray Computed Tomography. *Zootaxa* 53: 47–53. <https://www.scopus.com/record/display.url?eid=2-s2.0-35748974880&origin=resultlist>
- Philippe H, Derelle R, Lopez P, Pick K, Borchellini C, Boury-Esnault N, Vacelet J, Renard E, Houliston E, Quéinnec E, Da Silva C, Wincker P, Le Guyader H, Leys S, Jackson DJ, Schreiber F, Erpenbeck D, Morgenstern B, Wörheide G, Manuel M (2009) Phylogenomics revives traditional views on deep animal relationships. *Current Biology* 19: 706–712. doi: 10.1016/j.cub.2009.02.052
- Pianykh OS (2012) *Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide.* Springer, Berlin, 417 pp. doi: 10.1007/978-3-642-10850-1
- Postnov A, de Clerck N, Sasov A, Van Dyck D (2002) 3D in-vivo X-ray microtomography of living snails. *Journal of Microscopy* 205: 201–204. doi: 10.1046/j.0022-2720.2001.00986.x
- Prohaska S, Hutanu A, Hege H-C (2004) Interactive Exploration of Large Remote Micro-CT Scans. In: *Proceedings of the Conference on Visualization '04, San Diego (USA), October 2001 IEEE Computer Society Washington, DC, 345–352.* <https://www.scopus.com/record/display.url?eid=2-s2.0-17044418003&origin=resultlist>
- Puce S, Pica D, Brun F, Mancini L, Bavestrello G (2012) Genus *Distichopora* (Cnidaria, Hydrozoa): from primary cyclosystem to adult pore organisation. *Coral Reefs* 31: 715–730. doi: 10.1007/s00338-012-0885-0
- Qiong X, Hengyong Y, Bennet J, Peng H, Zainon R, Doesburg R, Opie E A, Walsh M, Haiou S, Butler A, Butler P, Xuanqin M, Ge W (2012) Image Reconstruction for Hybrid True-Color Micro-CT. *IEEE transactions on biomedical engineering* 59: 1711–1719. doi: 10.1109/TBME.2012.2192119
- Quatrefages A (1866) *Histoire naturelle des Annelés marins et d'eau douce. Annélides et Géphyriens, Volume 1.* Librairie Encyclopédique de Roret, Paris, 588 pp.

- Quintarelli G, Bellocchi M, Geremia R (1973) On phosphotungstic acid staining. IV. Selectivity of the staining reaction. *Journal of Histochemistry & Cytochemistry* 21: 155–160. doi: 10.1177/21.2.155
- Rowe T, Frank LR (2011) The disappearing third dimension. *Science* 331: 712–714. doi: 10.1126/science.1202828
- Ruthensteiner B, Baumler N, Barnes DG (2010) Interactive 3D volume rendering in biomedical publications. *Micron* 41: 886.e1–886.e17. doi: 10.1016/j.micron.2010.03.010
- Ruthensteiner B, Hess M (2008) Embedding 3D models of biological specimens in PDF publications. *Microscopy Research and Technique* 71: 778–786. doi: 10.1002/jemt.20618
- San Martín G (2003) Annelida Polychaeta II. Syllidae. In: Ramos MA et al. (Eds) *Fauna Ibérica*, vol. 21. Museo Nacional de Ciencias Naturales, CSIC, Madrid, 554 pp.
- Smith V, Blagoderov V (2012) Bringing collections out of the dark. *ZooKeys* 209: 1–6. doi: 10.3897/zookeys.209.3699
- Sutton MD (2008) Tomographic techniques for the study of exceptionally preserved fossils. *Proceedings of the Royal Society of London, Series B* 275: 1587–1593. doi: 10.1098/rspb.2008.0263
- Sutton MD, Garwood RJ, Siveter DJ, Siveter DJ (2012) SPIERS and VAXML; A software toolkit for tomographic visualisation and a format for virtual specimen interchange. *Paleontologica electronica* 15: 1–15. <http://palaeo-electronica.org/content/issue-2-2012-technical-articles/226-virtual-palaeontology-toolkit>
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28: 2731–2739. doi: 10.1093/molbev/msr121
- Wang G, Yu H, De Man B (2008) An outlook on x-ray CT research and development. *Medical Physics* 35: 1051–1064. doi: 10.1118/1.2836950
- Westheide W (1997) The direction of evolution within the Polychaeta. *Journal of Natural History* 31: 1–15. doi: 10.1080/00222939700770011
- Wheeler QD, Raven PH, Wilson EO (2004) Taxonomy: impediment or expedient? *Science* 303: 285. doi: 10.1126/science.303.5656.285
- Wheeler Q, Thierry B, Coddington J, Gostony T, Hamilton A, Larimer R, Polaszek A, Schauff M, Solis A (2012) Nomenclatural benchmarking: the roles of digital typification and telemicroscopy. *ZooKeys* 209: 193–202. doi: 10.3897/zookeys.209.3486
- Wilhelm G, Handschuh S, Plant J, Nemeschkal HL (2011) Sexual dimorphism in head structures of the weevil *Rhopalapion longirostre* (Olivier 1807) (Coleoptera: Curculionidae): a response to ecological demands of egg deposition. *Biological Journal of the Linnean Society* 104: 642–660. doi: 10.1111/j.1095-8312.2011.01751.x
- Wirkner CS, Prendini L (2007) Comparative Morphology of the Hemolymph Vascular System in Scorpions—A Survey Using Corrosion Casting, MicroCT, and 3D-Reconstruction. *Journal of Morphology* 413: 401–413. doi: 10.1002/jmor.10512
- Wolff S (1967) Radiation Genetics. *Annual Review of Genetics* 1: 221–244. doi: 10.1146/annurev.ge.01.120167.001253

- Zanol J, Bettoso N (2006) Identity of *Eunice roussaei* (Eunicidae: Polychaeta: Annelida) from the Adriatic and Mediterranean Seas. *Journal of the Marine Biological Association of the United Kingdom* 86: 1017–1024. doi: 10.1017/S0025315406013993
- Zanol J, Halanych KM, Struck TH, Fauchald K (2010) Phylogeny of the bristle worm family Eunicidae (Eunicida, Annelida) and the phylogenetic utility of noncongruent 16S, COI and 18S in combined analyses. *Molecular Phylogenetics and Evolution* 55: 660–676. doi: 10.1016/j.ympev.2009.12.024
- Ziegler A, Faber C, Mueller S, Bartolomaeus T (2008) Systematic comparison and reconstruction of sea urchin (Echinoidea) internal anatomy: a novel approach using magnetic resonance imaging. *BMC Biology* 6: 33. doi: 10.1186/1741-7007-6-33
- Ziegler A, Kunth M, Mueller S, Bock C, Pohmann R, Schröder L, Faber C, Giribet G (2011a) Application of magnetic resonance imaging in zoology. *Zoomorphology* 130: 227–254. doi: 10.1007/s00435-011-0138-8
- Ziegler A, Mietchen D, Faber C, von Hausen W, Schöbel C, Sellerer M, Ziegler A (2011b) Effectively incorporating selected multimedia content into medical publications. *BMC Medicine* 9: 17. doi: 10.1186/1741-7015-9-17
- Ziegler A, Ogurreck M, Steinke T, Beckmann F, Prohaska S, Ziegler A (2010) Opportunities and challenges for digital morphology. *Biology Direct* 5: 45. doi: 10.1186/1745-6150-5-45
- Zimmermann D, Randolph S, Metscher BD, Aspöck U (2011) The function and phylogenetic implications of the tentorium in adult Neuroptera (Insecta). *Arthropod Structure & Development* 40: 571–582. doi: 10.1016/j.asd.2011.06.003

Chapter 4

Faulwetter S, Markantonatou V, Pavloudi C, Papageorgiou N, Keklikoglou K, Chatzinikolaou E, Pafilis E, Chatzigeorgiou G, Vasileiadou K, Dailianis T, Fanini L, Koulouri P, Arvanitidis, C

(2014) *Polytraits*: A database on biological traits of marine polychaetes. *Biodiversity Data Journal* 2: e1024

doi: 10.3897/BDJ.2.e1024



Polytraits: A database on biological traits of marine polychaetes

Sarah Faulwetter^{†,‡}, Vasiliki Markantonatou^{§,||}, Christina Pavloudi^{†,||}, Nafsika Papageorgiou[‡], Kleoniki Keklikoglou[‡], Eva Chatzinikolaou[‡], Evangelos Pafilis[‡], Georgios Chatzigeorgiou[‡], Katerina Vasileiadou^{‡,#}, Thanos Dailianis[‡], Lucia Fanini[‡], Panayota Koulouri[‡], Christos Arvanitidis[‡]

[†] National and Kapodestrian University of Athens, Athens, Greece

[‡] Hellenic Centre for Marine Research, Heraklion, Crete, Greece

[§] Hellenic Centre for Marine Research, Heraklion, Greece

[|] Università Politecnica delle Marche, Ancona, Italy

^{||} Department of Biology, Faculty of Sciences, University of Ghent, 9000 Gent, Belgium, Department of Microbial Ecophysiology, Faculty of Biology, University of Bremen, 28359, Bremen, Germany

[#] Department of Biology, University of Patras, Rio, Patras, Greece

Corresponding author: Sarah Faulwetter (sarifa@hcmr.gr)

Academic editor: Cynthia Parr

Received: 18 Nov 2013 | Accepted: 16 Jan 2014 | Published: 17 Jan 2014

Citation: Faulwetter S, Markantonatou V, Pavloudi C, Papageorgiou N, Keklikoglou K, Chatzinikolaou E, Pafilis E, Chatzigeorgiou G, Vasileiadou K, Dailianis T, Fanini L, Koulouri P, Arvanitidis C (2014) *Polytraits*: A database on biological traits of marine polychaetes. Biodiversity Data Journal 2: e1024. doi: [10.3897/BDJ.2.e1024](https://doi.org/10.3897/BDJ.2.e1024)

Abstract

The study of ecosystem functioning – the role which organisms play in an ecosystem – is becoming increasingly important in marine ecological research. The functional structure of a community can be represented by a set of functional traits assigned to behavioural, reproductive and morphological characteristics. The collection of these traits from the literature is however a laborious and time-consuming process, and gaps of knowledge and restricted availability of literature are a common problem. Trait data are not yet readily being shared by research communities, and even if they are, a lack of trait data repositories and standards for data formats leads to the publication of trait information in forms which cannot be processed by computers. This paper describes *Polytraits* (<http://polytraits.lifewatchgreece.eu>), a database on biological traits of marine polychaetes (bristle worms, Polychaeta: Annelida). At present, the database contains almost 20,000 records on morphological, behavioural and reproductive characteristics of more than 1,000 marine polychaete species, all referenced by literature sources. All data can be freely accessed

through the project website in different ways and formats, both human-readable and machine-readable, and have been submitted to the [Encyclopedia of Life](#) for archival and integration with trait information from other sources.

Keywords

Polychaeta, biological traits, functional traits, morphology, reproduction, behaviour, larvae, life cycle, life history, database, literature

Introduction

Benthic organisms participate in a number of ecosystem-level processes, often described as “ecosystem functions”, which refer to any transformation process that occurs in an ecosystem (Cooper et al. 2008). The functional structure of a community can be represented by a set of functional traits assigned to behavioural, reproductive and morphological characteristics displayed by the observed species (Paganelli et al. 2012). Traits that affect resource use, feeding interactions, or habitat structure / availability are regarded as fundamentally important for ecosystem functioning (Bremner 2008), and the extent to which a species loss can threaten basic ecosystem processes depends on the functional richness (i.e., the number of functional groups) and evenness (i.e., the distribution of species across functional groups) in an ecosystem (Mouillot et al. 2005). Functional diversity is an important community property that can estimate the role organisms have in the ecosystem and can help to understand how the community reacts to environmental changes (Schleuter et al. 2010).

One approach to assess ecosystem functioning is to analyse species' biological traits which serve as a proxy for the functional characteristics of an assemblage (Bremner et al. 2003). This approach requires the selection of suitable traits that describe certain ecosystem processes and the assignment of species to these traits. However, this assignment is a laborious and time-consuming process that ideally requires collaborative work of a large team of people. Gaps of knowledge and information regarding species traits are a common problem (Tyler et al. 2012). Information may not be readily available since very often the relevant ecological and behavioural species studies are included in legacy literature which can be only found as rare paper copies in libraries around the world, or the biology of the species might not be known at all. Filling knowledge gaps requires both collection of additional data and development of statistical techniques for estimating missing trait values (Tyler et al. 2012). Lack of knowledge for certain traits in the literature is often surpassed by deriving available information for closely related species or even for species of the same family, since phylogenetically related species might have evolved similar environmental and ecological adaptations, thus leading to functional similarity (Usseglio-Polatera et al. 2000). However, the validity of this practice remains to be tested (Bevilacqua et al. 2012) and more accurate information on the evolutionary relationships between species is required (Tyler et al. 2012).

The organisation of the collected information in traits databases is an additional challenge. In most cases, trait data are not published together with the relevant manuscripts, or they are published as supplementary annexes in a format which is not machine-readable. The trend to share functional trait data over the internet, especially for the marine environment, is only recent. Examples of databases that provide trait information for marine species are the [Neogene Marine Biota of Tropical America](#) (NMITA) database, the [Biological Traits Information Catalogue](#) (BIOTIC) of the Marine Life Information Network and [SeaLifeBase](#). However, despite an increasing availability of these data through web-based databases, most of the data are not provided in a format that can be processed by computers. An obstacle to this is the lack of standardised data formats for describing trait data and the absence of ontologies (standardised and well-specified vocabulary of concepts and their interrelationships) with which trait information can be described. These are necessary prerequisites to integrate trait data into the semantic web – which will make them fully exploitable by both humans and computers. Furthermore, no public repository for traits data has existed to encourage authors to publish their data and make them re-usable. One initiative to address these problems is the [TraitBank initiative by the Encyclopedia of Life](#) (EOL), which will serve as a provider for aggregated species trait data in a machine-readable format.

This paper describes *Polytraits*, a database on biological traits of polychaetes (bristle worms, Polychaeta: Annelida). The database contains almost 20,000 records on morphological, behavioural and reproductive characteristics of more than 1,000 polychaete species (currently only marine species are covered). All data can be freely accessed through the [project website](#) in different ways and formats, both human-readable and machine-readable. Furthermore, the data are [available through EOL's TraitBank](#), seamlessly integrating the information with other relevant trait data.

General description

Purpose: The project was initially started as an in-house project of the Hellenic Centre for Marine Research, within the framework of the MSc thesis of V. Markantonatou. Traits were collected for an ecological study of polychaetes in Mediterranean lagoons. Since then, the database has been continuously expanded to serve the data needs for other analyses (e.g. as part of the PhD thesis of S. Faulwetter). The database and website are being maintained by the Institute for Marine Biology, Biotechnology and Aquaculture of the Hellenic Centre for Marine Research and will be supported by and constitute a part of the LifeWatch Greece infrastructure.

Project description

Title: *Polytraits*: A database of biological traits of polychaetes collected from the literature

Personel: Sarah Faulwetter (Concept development, database and web development, data collection and management, trait definitions), Christos Arvanitidis (Concept development, data collection, trait definitions), Vasiliki Markantonatou (Concept development, trait definitions, data collection), Nafsika Papageorgiou (trait definitions, data collection), Christina Pavlouti (trait definitions, data collection), Kleoniki Keklikoglou (data collection), Georgios Chatzigeorgiou (data collection), Katerina Vasileiadou (data collection), Evangelos Pafilis (data collection, web service development), Thanos Dailianis (data collection, website design), Lucia Fanini (data collection), Eva Chatzinikolaou (data collection), Panayota Koulouri (data collection)

Design description: To collect and disseminate biological trait data on polychaetes, a dedicated, web-based database has been created. It can be accessed at <http://polytraits.lifewatchgreece.eu>. The database features an entry interface (access only for registered users), and several options to access and export the data (see detailed documentation at the end of this manuscript). The data can be browsed furthermore a) through the [Encyclopedia of Life](http://www.scratchpads.eu) and b) via a Scratchpads (<http://www.scratchpads.eu>) installation, dedicated to polychaete research (<http://polychaetes.lifewatchgreece.eu>). This taxon-centric virtual research environment (Smith et al. 2009) allows browsing the taxonomic classification and retrieving various kinds of relevant information for each taxon, among which are also the collected biological traits (Fig. 1).

The screenshot shows the Scratchpads interface for the Polytraits database. The top navigation bar includes 'HOME', 'LITERATURE', 'MEDIA GALLERY', and 'POLYCHAETA'. A search bar is located in the top right. The left sidebar shows a taxonomic tree for POLYCHAETA, with Syllidae selected. The main content area displays the 'Syllidae' taxon page, including a table of biological traits.

Trait	Modality	Value	Verified
Age at first reproduction	5 - 2 months		yes
Age at first reproduction	2 - 6 months		yes
Body size (max)	0.2mm - 0.25 cm (200µm - 2.5 mm)		no
Body size (max)	0.25cm - 1 cm (2.5mm - 10 mm)		no
Body size (max)	1cm-5cm (10mm-50mm)		no
Body size (max)	2cm-5cm (20mm-50mm)		no
Body size (max)	2cm-5cm (20mm-50mm)		yes

Figure 1.
Screenshot of the Scratchpads interface to the *Polytraits* database.

The database contains 47 different traits describing the morphological, behavioural, reproductive and larval features as well as the environmental affinities of a taxon. Each trait has several sub-categories, so-called "modalities" (e.g. the trait "Mobility of the adult"

contains the modalities "crawler", "burrower", "swimmer" and "non-motile/semi-motile"). In total, 252 modalities are covered. For a detailed list of all traits and modalities, including their definition, see section "Traits Coverage" below. The expression of a trait in a taxon is coded in the database by declaring each modality as "present" or "absent". This allows to capture both ambiguous information in the literature concerning the expression of a trait in a taxon, as well as the presence of multiple modalities in a taxon (e.g. some taxa can both be crawlers and swimmers).

Each assignment of a modality (and its presence/absence value) to a taxon is connected to a literature reference. This assignment is mandatory, no data can be entered without specifying the source of the data; however, to capture undocumented knowledge, the option "Expert's judgement" can be specified. The database contains its own literature management functions and stores references, the user can, during data entry, simply choose the desired reference from a list (or add it, if it is not yet present). It is possible to enter more than one reference per taxon-modality-value entry, thus corroborating the assignment. Additionally, most records are accompanied by the quotation of the exact literature passage which has led to the coding of the information. This serves two purposes: a) other researchers can re-use the data and code them differently if the modalities present in the database are not suitable for a specific analysis, and b) since the interpretation of text and the coding of the data is often subjective the original text can serve as a quality control mechanism. However, since this feature has only been introduced recently, some legacy data exist in the database which cite the reference but do not quote the exact text excerpt; these gaps are gradually being filled in.

Sampling methods

Study extent: The dataset includes biological traits (morphological, behavioural, reproductive and larval traits) of polychaetes. Since the data were initially collected with a specific research question and dataset in mind, a large number of the species in the database are typical inhabitants of European lagoons. However, this core set of species has been expanded over time and now comprises species from a large number of habitats and from world-wide locations. At present, the database contains only marine species, but freshwater and terrestrial polychaetes will be covered in the future.

Sampling description: Data were collected from 950 different literature sources, the vast majority of which are scientific journals (Fig. 2a). However, on average most trait information is contained in books and review papers, whereas primary research papers often only contain information leading to the entry of one or very few records (Fig. 2b). Table 1 displays the references from which most of the trait data in the database originate. The system allows also to record experts' knowledge on taxa which is not documented in the literature; here the source of the data is specified as "Expert's judgement". Literature references were not "sampled" exhaustively by systematically coding all contained trait information, but since the data collection was purpose-driven, only the information needed

for the analysis at the time was coded. Table 1 therefore does not display the absolute trait information content for the listed references but for the data in the *Polytraits* database.

Table 1.

The most employed literature references (more than 100 taxon-modality records each) in the *Polytraits* database. The number of taxon-modality records supported by each reference is given, as well as the number of records for which the original text excerpt is quoted (see also section "Quality control").

A full list of all references used to collect data for the *Polytraits* database can be found in Suppl. material 3.

Reference	Nr. of records (total)	Nr. of records with text excerpt	Nr. of records without text excerpt
Rouse and Pleijel (2001)	5466	5466	0
Rouse and Pleijel (2006)	2252	2252	0
Hartmann-Schröder (1996)	1756	1755	1
Almeida et al. (2003)	1540	1540	0
Expert's Judgement	1131	763	368
MarLIN (2006)	530	324	206
Eckelbarger (2005)	502	497	5
Borja et al. (2000)	491	1	490
Fauchald and Jumars (1979)	452	327	125
Arvanitidis (1994)	450	220	230
San Martín (2003)	410	388	22
Wilson (1991)	346	324	22
Read (2006)	214	214	0
Vieitez et al. (2004)	203	201	2
Carson and Hentschel (2006)	201	34	167
Blake and Arnofsky (1999)	181	120	61
Tzetlin and Purschke (2005)	178	140	38
Holthe (1986)	139	117	22
Fauvel (1927)	126	23	103
Wolff (1973)	124	124	0
McHugh and Fong 2002	114	82	32
Kupriyanova et al. 2001	111	59	52

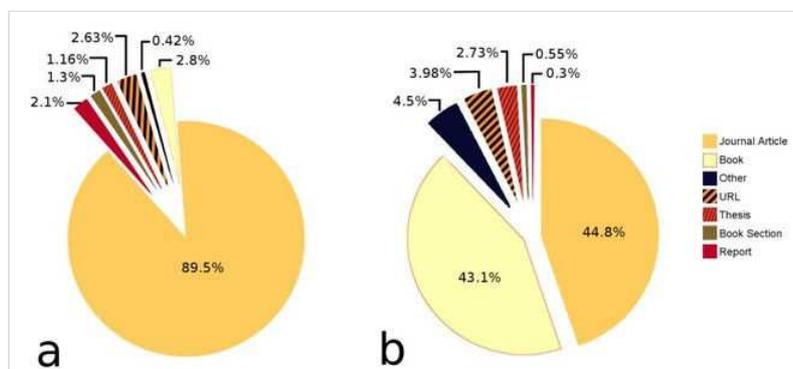


Figure 2.

a) Different types of literature sources stored in the *Polytraits* database; b) Origin of the information in the *Polytraits* database. Most information is contained in books and review papers, whereas on average, a single research paper contains little information. Graph based on data in Suppl. material 4 and Suppl. material 5.

To enter the trait information, a dedicated web-based entry interface had been developed, assuring the integrity of the entered data and preventing information from being entered without specifying the literature source. The information found in the literature was assigned to pre-defined trait categories (“modalities”, see below under Trait coverage), specifying either “presence” or “absence” of a modality in a taxon (or both, if this information was found in the literature).

The system allows data entry at different taxonomic levels (from subspecies to family). Data was always assigned to the most specific taxon possible (e.g. the information “*Eteone longa* is a predator” was assigned to *Eteone longa*; the information “Some species of *Eteone* are predators” was assigned to *Eteone*, the information “All Sabellidae are filter feeders” was assigned to all taxa in the family Sabellidae). Information for taxa that are at present considered as subjective synonyms was entered for each respective taxon involved, and not for the taxon presently considered as “valid”. This allows keeping information and taxonomy separate, and in case of a future resurrection of the synonymised taxon, the trait information does not have to be re-assigned. The structure of the database allows re-combining the information on synonyms during data export.

Quality control: For each record, the relevant text excerpt from the source literature was recorded in addition to the citation (in the original language of the text, to avoid the introduction of translation errors). This short excerpt – usually no more than one or few short sentences – allows the reader to understand what led to the assignment of the taxon to a specific trait category. It also provides a means for quality control and allows other researchers to re-use the information in different contexts. This is especially helpful if the research question in mind requires different trait categories from those that have been chosen in this database.

Since the option to record the original text passage had been introduced at a later stage during the project, a certain percentage of the entries still lacks this information (see details below under "Taxonomic Coverage" and "Traits coverage", as well as Table 1) which is being added gradually to improve the quality of the data. The quality status is indicated in the [Polychaetes Scratchpad](#) trait section with either a green tick or a red cross in the rightmost column (Fig. 1).

Step description: The data collection for the *Polytraits* database has mainly been purpose-driven, meaning that information has been collected for a specific set of taxa for which an analysis was then performed. However, if no information on a certain trait in a taxon could be found in the literature, information for synonyms, congeners and confamilials was likewise collected (from which data for the actual taxon in question can in some cases be derived). The data entry interface allows the registered user to enter new taxa and references. When a new taxon is entered, the taxonomic classification is automatically retrieved from the World Register of Marine Species and the taxon integrated into the local classification. All available information can be accessed and downloaded by the public through the [Polytraits website](#), the [Encyclopedia of Life](#), the [Polychaetes Scratchpad](#) or programmatically through a web service (see below under "Data resources" for further specifications of all access options). A schema of the data flow can be found in Fig. 3.

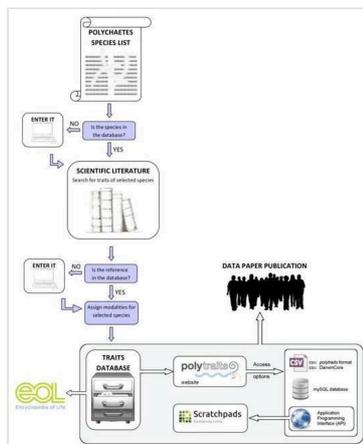


Figure 3.

Flowchart of the data entry and publication process of the biological traits data in the *Polytraits* database. Starting with a species list for which data are to be collected, the respective taxa are entered in the database by the user. For each taxon, the required information is gathered from the literature. Data are always connected to their literature reference in the database and often the original text passage is also stored. Once in the database, the data can be accessed through an application programming interface (API), which also serves the [Polychaetes Scratchpads](#), where data are displayed and can be browsed through a biological classification. Furthermore, data can be [downloaded](#) for offline use in various formats, and finally they can be browsed through the web portal of the [Encyclopedia of Life](#).

Geographic coverage

Description: A large number of taxa in the traits database are distributed in European waters. However, the database is constantly being expanded and by no means geographically restricted, so species from locations all over the world are also present in the database, albeit in lower numbers.

Taxonomic coverage

Description: At present, the database contains 19,632 taxon-modality-value entries for 1,133 species-level, genus-level and family-level taxa of polychaetes. Currently only marine taxa are covered, but future expansions of the database will cover terrestrial and freshwater species as well. No data have been collected for taxa higher than the rank of family. Different representations of a taxonomic name can be stored in the database (e.g. objective synonyms, misspellings); the total number of taxonomic names for these taxa accounts to 1,373 (Fig. 4). A full list of all taxa in the *Polytraits* database for which trait information is available can be found on the [website](#), including an overview of the amount of information available for each taxon. Several taxon-modality-value entries are supported by more than one literature source, so the total number of taxon-modality-value-reference entries amounts to 25,042. Updates to these numbers can be found on the [Polytraits website](#).

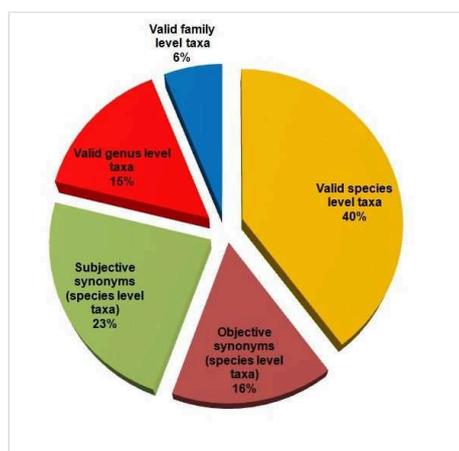


Figure 4.

Taxonomic rank and status of taxa in the *Polytraits* database. Graph based on data in Suppl. material 6.

Syllidae are the most species-rich family in the *Polytraits* database, followed by Spionidae and Nereididae (Fig. 5). Fig. 6 gives an overview of the granularity of the traits information available for each family: most information has been recorded at species level, though the

percentage of information recorded for higher taxa (genus, family) is higher in some families with a low number of taxa.

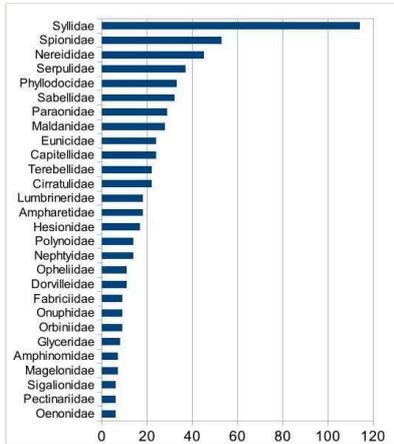


Figure 5.

Number of valid species contained in the *Polytraits* database, per family. Graph based on data in Suppl. material 7.

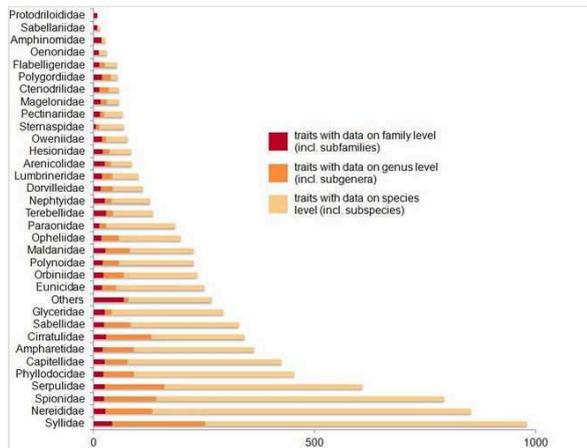


Figure 6.

Number of taxon-modality-value records on family, genus and species level, per family. Graph based on data in Suppl. material 8.

The families with the largest number of species in the database contain also the largest amount of information on their biological traits. Table 2 gives an overview of the number of records per family (for all taxa within that family). The number of records for which the original text excerpt is not recorded varies among families. This is a result of the way information was recorded: during data collections in the beginning of the project the option to record this data was not implemented and this information was only gradually added.

Taxa present in datasets for which traits data was collected recently have a higher percentage of recorded text excerpts.

Table 2.

Number of taxon-modality-literature records per family (for all taxa in each family), as well as the number and percentage of these records for which the original text excerpt from the literature has also been recorded.

Family	Total nr. of records	Nr. of entries with text excerpt	% of entries with text excerpt	% of entries without text excerpt
Nereididae	3168	3012	95.08	4.92
Syllidae	2860	2541	88.85	11.15
Spionidae	2190	1758	80.27	19.73
Serpulidae	1820	1529	84.01	15.99
Capitellidae	1469	1331	90.61	9.39
Phyllodocidae	1257	1189	94.59	5.41
Cirratulidae	1171	696	59.44	40.56
Ampharetidae	1103	987	89.48	10.52
Sabellidae	910	740	81.32	18.68
Glyceridae	818	740	90.46	9.54
Orbiniidae	710	532	74.93	25.07
Eunicidae	644	509	79.04	20.96
Polynoidae	615	562	91.38	8.62
Opheliidae	568	501	88.2	11.8
Paraonidae	561	453	80.75	19.25
Maldanidae	516	462	89.53	10.47
Terebellidae	500	259	51.8	48.2
Nephtyidae	480	422	87.92	12.08
Arenicolidae	455	121	26.59	73.41
Lumbrineridae	356	136	38.2	61.8
Oweniidae	339	152	44.84	55.16
Pectinariidae	283	257	90.81	9.19
Dorvilleidae	279	233	83.51	16.49
Hesionidae	279	176	63.08	36.92
Fabriciidae	278	234	84.17	15.83
Sternaspidae	197	191	96.95	3.05
Flabelligeridae	164	121	73.78	26.22
Magelonidae	153	63	41.18	58.82
Ctenodrilidae	152	113	74.34	25.66
Polygordiidae	130	76	58.46	41.54
Oeonidae	115	42	36.52	63.48
Trichobranchidae	71	58	81.69	18.31
Amphinomidae	60	55	91.67	8.33

Sabellariidae	33	31	93.94	6.06
Paralacydoniidae	32	32	100	0
Lacydoniidae	23	23	100	0
Chrysopetalidae	23	23	100	0
Protodriloididae	19	19	100	0
Pholoidae	18	18	100	0
Sphaerodoridae	16	16	100	0
Siboglinidae	10	10	100	0
Myzostomidae	9	9	100	0
Aberrantidae	8	8	100	0
Aphroditidae	8	7	87.5	12.5
Goniadidae	8	8	100	0
Onuphidae	8	8	100	0
Acrocirridae	7	7	100	0
Chaetopteridae	7	7	100	0
Pilargidae	7	7	100	0
Tomopteridae	7	7	100	0
Acoetidae	6	6	100	0
Aeolosomatidae	6	6	100	0
Alvinellidae	6	6	100	0
Fauveliopsidae	6	6	100	0
Nerillidae	6	6	100	0
Pisionidae	6	6	100	0
Poecilochaetidae	6	6	100	0
Parergodrilidae	6	6	100	0
Alciopidae	6	6	100	0
Cossuridae	5	5	100	0
Euphrosinidae	5	5	100	0
Scalibregmatidae	5	5	100	0
Sigalionidae	5	5	100	0
Nautiliniellidae	4	4	100	0
Questidae	4	4	100	0
Typhloscolecidae	4	4	100	0
Alciopidae	4	4	100	0
Eulepethidae	3	3	100	0
Histriobdellidae	3	3	100	0
Psammodrilidae	3	3	100	0
Spintheridae	3	3	100	0
Endomyzostomatidae	2	2	100	0
Lopadorhynchidae	2	2	100	0
Poeobiidae	2	2	100	0
Trochochaetidae	2	2	100	0
Apistobranchidae	1	1	100	0

Flotidae	1	1	100	0
Hartmaniellidae	1	1	100	0
Uncispionidae	1	1	100	0

Taxa included:

Rank	Scientific Name	Common Name
kingdom	Animalia	Animals
phylum	Annelida	Segmented worms
class	Polychaeta	Bristle worms

Traits coverage

The database contains 47 traits which are subdivided into 252 sub-categories (called "modalities"). They cover mainly reproductive and behavioural traits of both adult and larval stages, as well as information on environmental preferences and a few morphological traits.

Traits were chosen, defined and amended according to the needs of each analysis for which data was collected, thus they are a compilation of various sources. As a consequence several traits currently included have been recognised as inadequate to reflect polychaetes life histories and are likely to be changed (e.g. "Migration of adults", "Sociability", partly overlapping modalities of "Habitat" and "Physiographic feature"). Furthermore, the initial focus on European marine species has resulted in certain traits being defined appropriately for these taxa, but requiring an expansion of modalities if terrestrial and freshwater species, as well as species from other regions are included (e.g. expansion / refinement of salinity ranges, additional habitat terms). Future versions of this database will include a revised set of traits and the already existing data, but old versions will be available for download and changes to traits and their definitions will be properly documented.

No single suitable ontology (standardised vocabulary of concepts) exists to describe marine invertebrate traits. A sound definition of the concepts employed in the database and the identification of these concepts through Unique Resource Identifiers (URIs) allows the integration of the data into other trait data collections (such as the Encyclopedia of Life's TraitBank), therefore traits and modalities have been mapped, where possible, to existing ontology terms (e.g. the [Environment Ontology](#) or the [Animal Natural History and Life History ontology](#)).

Several traits and modalities are identical to those used by the [BIOTIC database](#) (MarLIN 2006), one of the most comprehensive databases on biological traits of marine organisms, to provide consistency across definitions and to ensure that data can be integrated more easily in the future. However, there is always a compromise required between trying to achieve comparability of the data and defining concepts as adequately as possible for a specific group of species, region and purpose. The approach used in this database is a

mixed model – employing existing ontology concepts where possible, but defining others specifically for the scope of this database. Future developments of the community concerning traits standards will therefore require continuous revisions of the traits and modalities used in the *Polytraits* database.

A full list of the traits, modalities and their definitions is given in Table 3. This table is a compact version of a more extensive documentation of the traits and modalities which can be found on the [Polytraits website](http://polytraits.lifewatchgreece.eu/terms), where additional information and references for each trait and modality are provided, as well as ontology mappings and identifiers for each trait and modality in form of a Unique Resource Identifier (URI).

Table 3.	
Definition of traits and their modalities in the <i>Polytraits</i> database. Please refer to the extended version of this table at http://polytraits.lifewatchgreece.eu/terms for additional descriptions and literature references, related terms and synonyms, as well as Unique Resource Identifiers (URIs) and ontology mappings for each trait and modality.	
Adult traits	
Body size (max)	A measurement of the longest dimension of a body, typically between two distinct ends of the body. In polychaetes, this is the length from the head to the pygidium without appendages like antennae or cirri. In the <i>Polytraits</i> database there are 7 different classes (modalities) for this trait. For the coding of modalities, the maximum body size that is reported in the literature for a species is chosen.
Modalities:	
< 2.5 mm	Maximum body size up to 2.5 mm.
2.5 mm – 10 mm	Maximum body size from 2.5 to 10 mm.
11 mm – 20 mm	Maximum body size from 11 to 20 mm.
21 mm – 50 mm	Maximum body size from 21 to 50 mm.
51 mm – 80 mm	Maximum body size from 51 to 80 mm.
81 mm – 100 mm	Maximum body size from 81 to 100 mm.
> 100 mm	Maximum body size more than 100 mm.
Complex species	A group of species which satisfy the biological definition of species, that is, they are reproductively isolated from each other, but they are not morphologically distinguishable (or at least are not readily or reliably distinguishable on a morphological basis) (Mayr and Ashlock 1991).
Modalities:	
yes	Complex species reported in the literature.
no	No complex species reported in the literature.
Depth zonation (benthos)	The depth at which an organism occurs. Commonly defined based on ecological features of the zonation.
Modalities:	

<i>supralittoral zone</i>	The zone of the shore immediately above the highest water level and subjected to wetting by spray or wave splash (Lincoln et al. 1998).
<i>littoral zone</i>	The area of the foreshore and seabed that is exposed to the air at low tide and submerged at high tide, i.e., the area between tide marks.
<i>sublittoral zone</i>	The zone of the shore immediately below the lowest water level and the edge of the continental shelf (ca. 200 m).
<i>bathyal zone</i>	The steep descent zone from 200 m to 4000 m depth.
<i>abyssal zone</i>	The zone between 4000 – 6000 m depth (Lincoln et al. 1998).
<i>hadal zone</i>	The sea floor deeper than 6000 m, such as that of the oceanic trenches.
Depth zonation (pelagic)	The depth at which an organism occurs in the water column. Commonly defined based on ecological features of the zonation.
<u>Modalities:</u>	
<i>epipelagic zone</i>	The zone of an ocean from the surface to 200 m where photosynthesis can occur, due to the penetration of light.
<i>mesopelagic zone</i>	Water column from the upper aphotic zone (ca. 200 m) to a depth of ca. 1000 m (MarLIN 2013).
<i>bathypelagic zone</i>	Water column from ca. 1000 m to a depth of ca. 2500 m (MarLIN 2013).
<i>abyssopelagic zone</i>	The zone of the ocean below the bathypelagic zone, with its lowest boundary at about 6000 m.
<i>hadalpelagic zone</i>	The zone of an ocean in oceanic trenches, lying between 6000 m and 10000 m.
Ecosystem engineering	Organisms can be considered as ecosystem engineers when they directly or indirectly modulate the availability of resources to other species, by causing physical state changes in biotic or abiotic materials. In so doing they modify, maintain and/or create habitats (Jones et al. 1994).
<u>Modalities:</u>	
<i>yes</i>	“Umbrella term”. Used to capture information that a species is an ecosystem engineer, without specifying the type of engineering.
<i>no</i>	“Umbrella term”. Used to capture information that a species is not an ecosystem engineer.
<i>biodiffusor</i>	Biodiffusors include organisms with activities that usually result in a constant and random local sediment biomixing over short distances (Kristensen et al. 2012).
<i>upward conveyor</i>	Upward conveyors are vertically oriented species that typically feed head-down at depth in the sediment. Vertically oriented head-down feeders actively select and ingest particles at the deeper sediments and egest these non-locally as faeces in the sediment surface (Kristensen et al. 2012).
<i>downward conveyor</i>	Downward conveyors exhibit a feeding strategy opposite to that of upward conveyors. Vertically oriented head-up feeders actively select and ingest particles at the surface and egest these non-locally as faeces in deeper sediment strata (Kristensen et al. 2012).
<i>regenerator</i>	Regenerators are excavators that dig and continuously maintain burrows in the sediment and by doing so they mechanically transfer sediment from depth to the surface.

<i>blind-ended ventilation</i>	Ventilation occurs when animals flush their burrows with overlying water for respiratory and feeding purposes. Blind-ended ventilation occurs when I-shaped burrows are flushed uni- or bidirectionally depending on the permeability of the sediment (Kristensen et al. 2012).
<i>open-ended ventilation</i>	In open-ended ventilation the burrows are U-shaped and can be flushed easily from one end to the other (Kristensen et al. 2012).
<i>habitat-building (reef-forming)</i>	Species which create structures which in turn form new habitats for other species.
Feeding structure	The feeding structures of the polychaetes vary, reflecting the diversity of feeding types. There are two major anatomical/morphological features involved in the polychaetes feeding: the pharynx and the feeding structures of the prostomium (e.g. palps) (Rouse and Pleijel 2006).
<u>Modalities:</u>	
<i>simple axial pharynx</i>	A sac-like pharynx relying on fluid pressure from the coelom for eversion (Rouse and Pleijel 2006).
<i>ventral buccal organ (simple)</i>	A variable set of folds, musculature and glands, present on the ventral side of many polychaetes, is usually referred to as a ventral pharynx and is the most common form in Polychaeta (Rouse and Pleijel 2006).
<i>ventral muscularpharynx</i>	The ventral and lateral walls of the buccal region are muscular and the lining is sclerotized into a varying number of eversible jaw pieces. The jaws are separated into a pair of ventral mandibles and two or more pairs of lateral maxillae (Rouse and Pleijel 2006).
<i>muscular axialpharynx</i>	The pharynx has thickened, strongly muscular walls and can be retracted into a sheath. In other cases the pharynx is partially retracted and partially inverted. The mouth proper is located at the tip of the pharynx when fully everted (Rouse and Pleijel 2006).
<i>buccal organ absent or occluded</i>	The buccal cavity lacks obvious differentiation of the wall and it is not eversible. In some species, if the buccal cavity is present at all, it is only a transient larval structure and becomes completely occluded (Rouse and Pleijel 2006).
<i>accessory feeding structures</i>	Other structures as palps, tentacles or a radiolar crown ("grooved palps").
Feeding type	The common diet of an organism that includes the food items which are enzymatically and behaviourally capable of using.
<u>Modalities:</u>	
<i>predator</i>	An organism that feeds by preying on other organisms, killing them for food (MarLIN 2013).
<i>suspension feeder</i>	Any organism which feeds on particulate organic matter, including plankton, suspended in the water column (MarLIN 2013).
<i>non-selective deposit feeder</i>	An organism that feeds on mud or sand and may show a little discrimination in the size or type of particles eaten. The sediment is ingested and any digestible organic material is assimilated as it passes through the alimentary canal.
<i>selective deposit feeder</i>	Some deposit feeders do not ingest sediment haphazardly but use their palps or buccal organs to sort organic material from the sediment prior to ingestion. The method of sorting varies according to the types of palps present.
<i>deposit feeder (selective or non-selective)</i>	"Umbrella term". Any organism which feeds on fragmented particulate organic matter from the substratum (MarLIN 2013). This modality should be filled in if nothing about the selectivity of the deposit feeding is known.

<i>omnivore</i>	Organisms which feed on a mixed diet including plant and animal material (MarLIN 2013).
<i>scavenger</i>	Any organism that actively feeds on dead animals.
<i>herbivore</i>	An animal that feeds on plants or algae, or parts of them.
Habitat type	The place in which an organism lives. It is defined for the marine environment according to geographical location, physiographic features as well as the physical and chemical environment (including salinity, wave exposure, strength of tidal streams, geology, biological zone, substratum, 'features' (e.g. crevices, overhangs, rockpools) and 'modifiers' (e.g. sand-scour, wave-surge, substratum mobility) (MarLIN 2013). The modalities of this trait might be expanded in the future and/or merged with the trait "Physiographic feature".
Modalities:	
<i>algae</i>	Macroalgae surfaces, such as <i>Laminaria</i> spp., or fucoids.
<i>biogenic reef</i>	Solid, massive structure which is created by accumulations of organisms, usually rising from the seabed, or at least clearly forming a substantial, discrete community or habitat which is very different from the surrounding seabed. The structure of the reef may be composed almost entirely of the reef building organism and its tubes or shells, or it may to some degree be composed of sediments, stones and shells bound together by the organisms (Holt et al. 1998).
<i>caves</i>	A hollow normally eroded in a cliff, with the penetration being greater than the width at the entrance (Sunamura 1992). Caves can also be formed by boulders (MarLIN 2013).
<i>crevices / fissures</i>	Crevices are narrow cracks in a hard substratum < 10 mm wide at its entrance, with the penetration being greater than the width at the entrance. Fissures are cracks in a hard substratum > 10 mm wide at its entrance, with the depth being greater than the width at the entrance (MarLIN 2013).
<i>maerl / coralligenous habitats</i>	A coralligenous habitat is defined by the presence of a bioherm of coralline algae grown at low irradiance levels and in relatively calm waters (Ballesteros 2006). Maerl denotes loose-lying, normally non-geniculate (i.e. not jointed), coralline red algae. Depending on the terminology used, maerl refers either to a class of rhodoliths, or may be considered distinct from rhodoliths in lacking a non-algal core. Maerl beds are composed of living or dead unattached corallines forming accumulations with or without terrigenous material (Birkett et al. 1998).
<i>other species</i>	Epibiont of other species.
<i>overhangs</i>	An overhanging part of a rock formation.
<i>rockpools</i>	A depression in the littoral zone of a rocky seashore, where, during low tide, seawater is left behind (MarLIN 2013).
<i>salt marsh</i>	A marsh whose water contains a considerable quantity of dissolved salts.
<i>seagrass</i>	Habitat associated with seagrass meadows communities. Seagrasses are flowering plants that are adapted to living fully submerged and rooted in estuarine and marine environments (MarLIN 2013).
<i>strandline</i>	A line on the shore composing debris deposited by a receding tide; commonly used to denote the line of debris at the level of extreme high water (MarLIN 2013).
<i>under boulders</i>	Under unattached rocks that can be very large (> 1024 mm), large (512 – 1024 mm) or small (256 – 512 mm) (MarLIN 2013).
<i>water column</i>	Pelagic habitat.

<i>soft sediments</i>	Deposits with a high water content (near or above the liquid limit), where the percolating skeleton is made of fine-grained soils (clay fraction above ~ 20%), with a high degree of saturation, and subjected to low effective confinement (Klein and Santamarina 2005).
Intra- and interspecific competition	The simultaneous demand by two or more organisms or populations or species for an essential common resource that is actually or potentially in limited supply or the detrimental interaction between two or more organisms or species seeking a common resource that is not limited (Eleftheriou 1997).
<u>Modalities:</u>	
<i>annelida (adults)</i>	Competition with other annelids that are in adult stage. The interaction can be between different organisms, populations or species.
<i>crustacea (adults)</i>	Competition with crustaceans that are in adult stage.
<i>annelida (larvae)</i>	Competition with other annelids that are in larval stage. The interaction can be between different organisms, populations or species.
<i>crustacea (larvae)</i>	Competition with crustaceans that are in larval stage.
<i>mollusca</i>	Competition with mollusks.
Lifespan	Maximum length of time that any particular organism can be expected to live.
<u>Modalities:</u>	
≤ 1 year	Life span shorter than a year.
1 – 3 years	Life span between 1 and 3 years.
3 – 5 years	Life span between 3 and 5 years.
≥ 5 years	Life span more than 5 years.
Migrations of adult	Movement of an organism or group from one habitat or location to another (MarLIN 2013). This trait is poorly defined for invertebrates. Within the context of this database it will likely be redefined or become obsolete.
<u>Modalities:</u>	
<i>yes</i>	“Umbrella term”. Used to capture information that a species in its adult stage is migratory.
<i>no</i>	“Umbrella term”. Used to capture information that a species in its adult stage is non-migratory and remains within the same area.
Mobility of adult	The capability of an organism to move spontaneously and freely (MarLIN 2013).
<u>Modalities:</u>	
<i>crawler</i>	An organism that moves along on the substratum via movements of its legs, appendages (e.g. parapodia and chaetae) or muscles (MarLIN 2013).
<i>burrower</i>	An organism that lives or moves in a burrow in soft sediments.
<i>swimmer</i>	An organism that swims through the water column via movements of its fins, legs or appendages, via undulatory movements of the body or via jet propulsion; includes pelagic phases during reproduction (swarming at the surface) (MarLIN 2013).
<i>non-motile / semi-motile</i>	Permanently attached to a substratum (non-motile) or capable of movement across (or through) it (semi-motile) (MarLIN 2013).

Physiographic feature	The general physical characteristics of the marine environment in which an organism lives (MarLIN 2013). The modalities of this trait might be expanded in the future and /or merged with the trait "Habitat".
<u>Modalities:</u>	
<i>open coast</i>	Any part of the coast not within a marine inlet, strait or lagoon, including offshore rocks and small islands. This includes MNCR types; Linear coast, Islands / Rocks and Semi-enclosed coast.
<i>offshore seabed</i>	Seabed beyond three miles (5 km) from the shore.
<i>strait</i>	Strait is a narrow channel of water that connects two larger bodies of water, and thus lies between two land masses.
<i>fjord</i>	Fjord is a long and narrow sea inlet with high steeply sloped walled sides. A fjord is a landform created during a period of glaciation. Includes also sea lochs.
<i>ria</i>	Ria is a submergent coastal landform where sea levels rise either in relation to the land or as a result of eustatic sea level change; where the global sea levels rise or isostatic sea level change; where the land sinks. When this happens valleys which were previously at sea level become submerged. Includes also voes.
<i>estuary</i>	A semi-enclosed coastal body of water with one or more rivers or streams flowing into it, and with a free connection to the open sea.
<i>enclosed coast / embayment</i>	An area of water bordered by land on three sides. Includes also harbours and marinas.
<i>lagoon</i>	Enclosed bodies of water separated or partially separated from the sea by shingle, sand or sometimes rock and with a restricted exchange of water with the sea, yielding varying salinity regimes.
<i>hydrothermal vents</i>	A marine hydrothermal vent is a marine benthic feature where heat generated due to tectonic activity, either at divergent plate boundaries or convergent ocean plates where back-arc spreading occurs, is released or 'vented' to the surface. The resultant high temperature water jets are laden with dissolved metals and minerals.
Predated by	Organism categories that feed by preying on the present species. The modalities of this trait might be expanded in the future.
<u>Modalities:</u>	
<i>annelids</i>	Prey for other annelids.
<i>crustaceans</i>	Prey for crustacean species.
<i>fish</i>	Prey for fish species.
<i>birds</i>	Prey for bird species.
<i>mollusks</i>	Prey for mollusks.
<i>echinoderms</i>	Prey for echinoderm species.
Sociability	"Umbrella term". Used to capture information that an organism, population or species is living alone or interacting with others forming groups/communities or colonies (through asexual reproduction). This term and its modalities are poorly defined and will likely be re-defined within the context of this database.
<u>Modalities:</u>	
<i>algae</i>	Species that interact/live with algae.

<i>seagrasses</i>	Species that interact/live with seagrass meadows.
<i>annelids</i>	Species that interact/live with annelids.
<i>bacteria</i>	Species that interact/live with bacteria.
<i>crustaceans</i>	Species that interact/live with crustaceans.
<i>fish</i>	Species that interact/live with fish.
<i>mollusks</i>	Species that interact/live with mollusks.
<i>nematodes</i>	Species that interact/live with nematodes.
<i>echinoderms</i>	Species that interact/live with echinoderms.
<i>cnidarians</i>	Species that interact/live with cnidarians.
<i>poriferans</i>	Species that interact/live with poriferans.
<i>branchiostomids</i>	Species that interact/live with branchiostomids.
<i>bryozoans</i>	Species that interact/live with bryozoans.
<i>entoproctans</i>	Species that interact/live with entoproctans.
Substrate type	The surface on which an organism lives. The substrate may simply provide structural support, or may provide nutrients (MarLIN 2013).
<u>Modalities:</u>	
<i>bedrock</i>	Any stable hard substratum not separated into boulders or smaller sediment units.
<i>large to very large boulders</i>	Unattached rock, of large (512 – 1024 mm) or very large (> 1024 mm) size (MarLIN 2013).
<i>small boulders</i>	Unattached rock, of small (256 – 512 mm) size (MarLIN 2013).
<i>cobbles</i>	Sediment characterised by an average particle diameter between 64 and 256 mm.
<i>pebbles</i>	Sediment characterised by an average particle diameter between 4 and 64 mm.
<i>gravel</i>	An environmental material which is composed of pieces of rock that are at least two millimeters (2 mm) in its largest dimension and no more than 75 millimeters.
<i>sandy gravel</i>	50 – 80% gravel; 20 – 50% sand.
<i>muddy gravel</i>	50 – 80% gravel; 20 – 50% mud.
<i>muddy sandy gravel</i>	50 – 80% gravel; 20 – 50% mud and sand.
<i>coarse clean sand</i>	Sediment particles diameter between 0.5 – 4 mm; the sand fraction is > 80%.
<i>fine clean sand</i>	Sediment particles diameter between 0.063 – 0.5 mm; the sand fraction is > 80%.
<i>gravelly sand</i>	50 – 80% sand; 20 – 50% gravel.
<i>muddy gravelly sand</i>	50 – 80% sand; 20 – 50% mud and sand.
<i>muddy sand</i>	50 – 80% sand; 20 – 50% mud.
<i>sandy mud</i>	50 – 80% mud; 20 – 50% sand.
<i>sandy gravelly mud</i>	50 – 80% mud; 20 – 50% sand and gravel.
<i>gravelly mud</i>	50 – 80% mud; 20 – 50% gravel.
<i>mud</i>	Fine particles of silt and/or clay < 0.063 mm; the silt/clay fraction is > 80% (MarLIN 2013).
<i>silt</i>	Sediment characterised by an average particle diameter between 3.9 and 63 micrometers.

<i>clay</i>	Sediment characterised by an average particle diameter between 1 and 3.9 micrometers.
<i>mixed</i>	Mixtures of a variety of sediment types composed of pebble/gravel/sand/mud. This category includes muddy gravels, muddy sandy gravels, gravelly muds, and muddy gravelly sands.
<i>artificial</i>	E.g. wood, metal or concrete structures.
Survival salinity	The range of salinity in which an organism is capable to survive and grow (MarLIN 2013). The modalities of this trait will be refined in the future to capture also hypersaline and freshwater species.
<u>Modalities:</u>	
<i>full salinity</i>	The capability of an organism to live in environments of average marine water salinity (30 – 40 ‰).
<i>variable salinity</i>	The capability of an organism to live in environments of variable salinity (18 – 40 ‰).
<i>reduced salinity</i>	The capability of an organism to live in brackish water having a wide range of salinity between 18 ‰ and 30 ‰.
<i>low salinity</i>	The capability of an organism to live in brackish water with low salinity (< 18 ‰).
Survival temperature	The range of temperature in which an organism is capable to survive and grow.
<u>Modalities:</u>	
<i>cold waters</i>	The capability of an organism to live in cold water environments (< 0 – 10 °C).
<i>warm / temperate / subtropical waters</i>	The capability of an organism to live in environments of average temperatures (10 – 25 °C).
<i>tropical waters</i>	The capability of an organism to live in warm water environments (> 25 °C).
Tolerance (AMBI index)	The sensitivity of an organism to organic enrichment, classified through the AMBI index (Borja et al. 2000).
<u>Modalities:</u>	
<i>group I</i>	Species very sensitive to organic enrichment and present under unpolluted conditions (initial state).
<i>group II</i>	Species indifferent to enrichment, always present in low densities with non-significant variations with time (from initial state to slightly unbalanced condition).
<i>group III</i>	Species tolerant to excess organic matter enrichment. These species may occur under normal conditions, but their populations are stimulated by organic enrichment (slightly unbalanced condition).
<i>group IV</i>	Second-order opportunistic species (slightly to pronouncedly unbalanced condition).
<i>group V</i>	First-order opportunistic species (pronouncedly unbalanced condition).
Tube/burrow material	Materials used for the construction of an organism's tube or burrow (if present).
<u>Modalities:</u>	
<i>clay</i>	Tubes/burrows constructed of clay, a group of hydrous aluminium phyllosilicate minerals that are typically less than 3.9 micrometres in diameter.
<i>gravel</i>	Tubes/burrows constructed of gravel, an environmental material which is composed of pieces of rock that are at least two millimeters (2 mm) in its largest dimension and no more than 75 millimeters.

<i>sand</i>	Tubes/burrows constructed of sand, a naturally occurring granular material composed of finely divided rock and mineral particles.
<i>mud</i>	Tubes/burrows constructed of mud, a liquid or semi-liquid mixture of water and fine particles of silt and/or clay < 0.063 mm; the silt/clay fraction is > 80% (MarLIN 2013).
<i>biogenic detritus</i>	Tubes/burrows constructed of dead skeleton materials found in the environment (e.g. shells, algal parts).
<i>secretions</i>	Tubes/burrows constructed of bodily secretions, usually mucus.
<i>calcium carbonate</i>	Tubes constructed of calcium carbonate.
Typically feeds on	The type of food an organism prefers. The modalities of this trait might be expanded in the future.
<u>Modalities:</u>	
<i>algae</i>	Algae as food source.
<i>bacteria</i>	Bacteria as food source.
<i>annelids</i>	Annelida as food source
<i>ciliates</i>	Ciliates as food source.
<i>crustaceans</i>	Crustaceans as food source.
<i>diatoms</i>	Diatoms as food source.
<i>flagellates</i>	Flagellates as food source.
<i>foraminiferans</i>	Foraminiferans as food source.
<i>mollusks</i>	Mollusks as food source.
<i>detritus</i>	Particles of organic material from dead and decomposing organisms as food source.
<i>sediment</i>	Unselective ingestion of sediment.
<i>fish</i>	Fish, incl. their larvae, as food source.
<i>ascidians</i>	Ascidians, incl. their larvae, as food source.
<i>echinoderms</i>	Echinoderms, incl. their larvae, as food source.
<i>cnidarians</i>	Cnidarians as food source.
Larval stage traits	
Habitat type of settlement / early development	Habitat type of the larval settlement and early development after metamorphosis.
<u>Modalities:</u>	
<i>algae</i>	Macroalgae surfaces, such as <i>Laminaria</i> spp., or fucoids.
<i>biogenic reef</i>	Solid, massive structure which is created by accumulations of organisms, usually rising from the seabed, or at least clearly forming a substantial, discrete community or habitat which is very different from the surrounding seabed. The structure of the reef may be composed almost entirely of the reef building organism and its tubes or shells, or it may to some degree be composed of sediments, stones and shells bound together by the organisms (Holt et al. 1998).
<i>caves</i>	A hollow normally eroded in a cliff, with the penetration being greater than the width at the entrance (Sunamura 1992). Caves can also be formed by boulders (MarLIN 2013).

<i>crevices / fissures</i>	Crevices are narrow cracks in a hard substratum < 10 mm wide at its entrance, with the penetration being greater than the width at the entrance. Fissures are cracks in a hard substratum > 10 mm wide at its entrance, with the depth being greater than the width at the entrance (MarLIN 2013).
<i>maerl / coralligenous habitats</i>	A coralligenous habitat is defined by the presence of a bioherm of coralline algae grown at low irradiance levels and in relatively calm waters (Ballesteros 2006). Maerl denotes loose-lying, normally non-geniculate (i.e. not jointed), coralline red algae. Depending on the terminology used, maerl refers either to a class of rhodoliths, or may be considered distinct from rhodoliths in lacking a non-algal core. Maerl beds are composed of living or dead unattached corallines forming accumulations with or without terrigenous material (Birkett et al. 1998).
<i>other species</i>	Epibiont of other species.
<i>overhangs</i>	An overhanging part of a rock formation.
<i>rockpools</i>	A depression in the littoral zone of a rocky seashore, where, during low tide, seawater is left behind (MarLIN 2013).
<i>salt marsh</i>	A marsh whose water contains a considerable quantity of dissolved salts.
<i>seagrass</i>	Habitat associated with seagrass meadows communities. Seagrasses are flowering plants that are adapted to living fully submerged and rooted in estuarine and marine environments (MarLIN 2013).
<i>strandline</i>	A line on the shore composing debris deposited by a receding tide; commonly used to denote the line of debris at the level of Extreme High Water (MarLIN 2013).
<i>under boulders</i>	Under unattached rocks that can be very large (> 1024 mm), large (512 – 1024 mm) or small (256 – 512 mm) (MarLIN 2013).
<i>water column</i>	Pelagic habitat.
<i>soft sediments</i>	Deposits with a high water content (near or above the liquid limit), where the percolating skeleton is made of fine-grained soils (clay fraction above ~ 20%), with a high degree of saturation, and subjected to low effective confinement (Klein and Santamarina 2005).
Juvenile mobility	The capability of a juvenile to move spontaneously and freely.
<u>Modalities:</u>	
<i>crawler</i>	An organism that moves along on the substratum via movements of its legs, appendages (e.g. parapodia and chaetae) or muscles (MarLIN 2013).
<i>burrower</i>	An organism that lives or moves in a burrow in soft sediments.
<i>swimmer</i>	An organism that moves through the water column via movements of its fins, legs or appendages, via undulatory movements of the body or via jet propulsion; includes pelagic phases during reproduction (swarming at the surface) (MarLIN 2013).
<i>non-motile / semi-motile</i>	Permanently attached to a substratum (non-motile) or capable of moving across (or through) it (semi-motile) (MarLIN 2013).
Larval development	The mode of development from the larval to the adult stage.
<u>Modalities:</u>	
<i>direct development</i>	There are no intermediate larval stage(s) or postembryonic metamorphoses of any kind. Embryonic development culminates in the hatching or birth of a fully formed, albeit miniature adult (Hall and Olson 2003).
<i>indirect development</i>	One or more successive, free-living larval stages intervene between embryo and adult, with a more-or-less abrupt transition, or metamorphosis, between the last larval stage and the adult (Hall and Olson 2003).

Larval feeding type	The existence of two distinct larval types, feeding and non-feeding, has established the current paradigm for larval ecology. Feeding larvae are those that can capture and use exogenous food, whereas non-feeding larvae are those that cannot capture or use exogenous food (McEdward 1995).
<u>Modalities:</u>	
<i>planktotrophic</i>	A larval development strategy in which small eggs are converted into larger juveniles by means of larval feeding and growth (Levin and Bridges 1995).
<i>maternally derived nutrition</i>	“Umbrella term” describing the maternal sources of nutrition and including the terms lecithotrophy, adelphophagy, and translocation of nutrients.
Larval mode of development	Larvae development in the water column or on/in soft- or hard-bottom substrates
<u>Modalities:</u>	
<i>benthic</i>	Development on or near the bottom of a water body.
<i>pelagic</i>	Development in the water column.
Location of parental care	Defines the location of the parental care (if provided), either near the body of the parent or at a distance from it.
<u>Modalities:</u>	
<i>outside microenvironment of the parent</i>	Parental care is provided through e.g. protective structures, but not on the body of the parent or in its immediate living environment (e.g. in a burrow, tube or nest).
<i>within microenvironment of the parent</i>	Parental care is provided either on the body of the parent or in its immediate living environment (e.g. in a burrow, tube or nest).
Metamorphosis amount	Generally, any anatomical remodelling between opposing life-history periods, i.e. between the larval and the adult stage, can be considered as a form of metamorphosis (Nielsen 2000, Nielsen 2009). These changes can be rapid and cataclysmic, or can proceed gradually, depending on the particular developmental basis for the juvenile body plan within the body of the larva (Bishop et al. 2006).
<u>Modalities:</u>	
<i>catastrophic</i>	The metamorphosis is accompanied by massive internal change coupled with catastrophic destruction of the larval tissues. Huge chunks of the larval body, its tissues and organs, are digested away and reabsorbed, or simply discarded (Ryan 2011).
<i>non-catastrophic</i>	The adult develops from the juvenile through a process of extension and differential growth, including different larval stages but without a drastic change of the body plan.
Parental care / Brood protection	Any parental trait that enhances the fitness of a parent’s offspring after the offspring is released from the female body (Smiseth et al. 2012). Viviparity and other forms of lecithotrophy are excluded here from this definition and not considered as forms of parental care.
<u>Modalities:</u>	
<i>yes</i>	“Umbrella term”. Used to capture information that a species provides parental care to its offspring.
<i>no</i>	Used to capture information that a species does not provide parental care to its offspring beyond supplying them with a small package of yolk that serves as an initial source of nutrition until the offspring are fully capable of feeding for themselves (Smiseth et al. 2012).

Substrate type of settlement	Settlement is defined as the behavioural performance when pelagic larvae descend from the plankton to the benthos, and move upon the substratum with or without attaching to it. Settlement is reversible: a larva can swim up again from the substrate to resettle at another location (Qian and Dahms 2006). The surface on which larvae choose to settle is defined as the substrate of settlement.
<u>Modalities:</u>	
<i>hard substrates</i>	“Umbrella term”. Used to capture information that larvae choose some type of hard substrate for their settlement.
<i>sand</i>	Particles defined in three size categories: very coarse sand and granules (1 – 4 mm); medium and coarse sand (0.25 – 1 mm); very fine and fine sand (0.063 – 0.25 mm) (MarLIN 2013).
<i>mud</i>	Fine particles of silt and/or clay, < 0.063 mm diameter; the silt/clay fraction is > 80% (MarLIN 2013).
<i>clay</i>	Sediment characterised by an average particle diameter between 1 and 3.9 micrometers.
<i>silt</i>	Sediment characterised by an average particle diameter between 3.9 and 63 micrometers.
<i>gravel</i>	An environmental material which is composed of pieces of rock that are at least two millimeters (2 mm) in its largest dimension and no more than 75 millimeters.
<i>pebbles</i>	Sediment characterised by an average particle diameter between 4 and 64 mm.
<i>cobbles</i>	Sediment characterised by an average particle diameter between 64 and 256 mm.
<i>boulders</i>	Sediment characterised by an average particle diameter greater than 256 mm.
<i>bacterial / organic biofilm</i>	A complex aggregation of microorganisms marked by the excretion of a protective and adhesive matrix; usually adhering to a substratum.
Reproduction traits	
Age at reproductive maturity	Beginning of the first sexual reproductive cycle. Defined as the first ripening of gametes.
<u>Modalities:</u>	
<i>≤ 2 months</i>	Reproductive maturity reached at an age younger than 2 months.
<i>2 – 6 months</i>	Reproductive maturity reached at an age between 2 to 6 months.
<i>6 months–1 year</i>	Reproductive maturity reached at an age between 6 months to a year.
<i>1 – 2 years</i>	Reproductive maturity reached at an age between 1 to 2 years.
<i>2 – 3 years</i>	Reproductive maturity reached at an age between 2 to 3 years.
<i>3 – 4 years</i>	Reproductive maturity reached at an age between 3 to 4 years.
<i>≥ 4 years</i>	Reproductive maturity reached at an age more than 4 years.
Developmental mechanism	
Developmental mechanism	The mechanism of the development of the embryo(s), inside or outside of the parental organism.
<u>Modalities:</u>	
<i>oviparous</i>	Reproduction in which eggs are released by the female; development of offspring occurs outside the mother’s body.
<i>viviparous</i>	Reproduction in which fertilization and development take place within the female body and the developing embryo derives nourishment from the female.

Egg size	Diameter of the eggs spawned by an organism.
<u>Modalities:</u>	
<i>0–100 µm</i>	Egg diameter up to 100 µm.
<i>100–200 µm</i>	Egg diameter from 100 µm to 200 µm.
<i>> 200 µm</i>	Egg diameter larger than 200 µm.
Epitoky	Form of reproduction of marine polychaetes in which the new individual arises by modification and separation from the posterior end of the worm in order to leave the bottom and reproduce (MarLIN 2013).
<u>Modalities:</u>	
<i>yes</i>	The organism undergoes epitokous metamorphosis.
<i>no</i>	The organism does not undergo epitokous metamorphosis.
Factors triggering reproduction	Factors that can initiate or enhance reproduction.
<u>Modalities:</u>	
<i>lunar cycle</i>	Reproduction which is timed to particular phases of the lunar cycle (or the semilunar cycle of spring and neap tides) (Dorresteijn and Westheide 1999).
<i>pheromones / hormones</i>	Spawning as a result of a pheromonal interaction between swarming males and females. Hormonal factors may be involved not only in the timing of reproduction but also in sexual differentiation (Dorresteijn and Westheide 1999).
<i>photoperiod</i>	Reproduction which is timed to a particular daylight length (Dorresteijn and Westheide 1999).
<i>temperature</i>	Reproduction which is controlled by changes in water temperature. In some species, a certain temperature value is a prerequisite for reproduction to occur (Dorresteijn and Westheide 1999).
<i>salinity</i>	Reproduction which is stimulated by changes in salinity (George 1966).
Fecundity	The potential reproductive capacity of an organism or population, measured by the number of gametes (eggs) or asexual propagules (MarLIN 2013).
<u>Modalities:</u>	
<i>1 – 50</i>	Up to 50 eggs per female and reproductive event.
<i>50 – 500</i>	From 50 to 500 eggs per female and reproductive event.
<i>500 – 2500</i>	From 500 to 2500 eggs per female and reproductive event.
<i>2500 – 10000</i>	From 2500 to 10000 eggs per female and reproductive event.
<i>10000 – 20000</i>	From 10000 to 20000 eggs per female and reproductive event.
<i>20000 – 100000</i>	From 20000 to 100000 eggs per female and reproductive event.
<i>> 100000</i>	More than 100000 eggs per female and reproductive event.
Fertilization	Location of the union of male and female gametes.
<u>Modalities:</u>	
<i>internal</i>	Fertilization takes place within the female's body.
<i>external (broadcast spawner)</i>	A method of reproduction during which the gametes (egg and sperm) unite outside the body.

<i>external (pseudocopulation)</i>	A form of external fertilization in which the partners are in close contact (Rouse and Pleijel 2006).
Mode of reproduction	The production by an organism of new individuals that contain some portion of genetic material inherited from that organism.
<u>Modalities:</u>	
<i>gonochoristic</i>	Having separate sexes throughout life (MarLIN 2013).
<i>simultaneous hermaphrodite</i>	Condition of hermaphroditic animals (and plants) in which the reproductive organs of both sexes are present and functional at the same time.
<i>sequential hermaphrodite</i>	Sequential hermaphrodites are born as one sex, but can later change into the opposite sex. Can be subdivided into protandrous and protogynous hermaphroditism.
<i>asexual reproduction</i>	Reproduction that is not sexual; that is, reproduction that does not include recombining the genotypes of two parents. Includes all different types of asexual reproduction (budding; parthenogenesis etc).
Pattern of oogenesis	Process of germ cell development in the female from the primordial germ cells through oogonia to the mature haploid ova. In polychaetes, two patterns have been identified: intraovarian and extraovarian (Eckelbarger 1983).
<u>Modalities:</u>	
<i>intraovarian</i>	Occurs when oocytes are retained by the ovary until most or all of oogenesis (and vitellogenesis) is completed. Ovaries are usually large, structurally complex, and persistent throughout the sexual phase of the female (Rouse and Pleijel 2006).
<i>extraovarian</i>	Occurs when small, previtellogenic oocytes are released from the ovary and complete vitellogenesis in the fluid-filled coelom. Ovaries are generally small, simple and sometimes have a transient nature (Rouse and Pleijel 2006).
Population sex ratio	The ratio of male to female (or vice versa) in a population.
<u>Modalities:</u>	
<i>1:1</i>	The ratio of female to male in the population is 1 to 1.
<i>female > male</i>	The number of females is higher in a population.
<i>female < male</i>	The number of males is higher in a population.
Reproduction strategy of the individual	Number of times an individual can reproduce during its lifetime.
<u>Modalities:</u>	
<i>iteroparous</i>	Breeding several times per lifetime.
<i>semelparous</i>	Organisms that have only one brood during their life time and then the parent usually dies.
Reproduction temperature	Water temperature that initiates or enhances reproduction.
<u>Modalities:</u>	
<i>cold water</i>	Reproduction in cold water environments (< 0 – 10 °C).
<i>warm / temperate / subtropical waters</i>	Reproduction in environments of average temperatures (10 – 25 °C).
<i>tropical waters</i>	Reproduction in warm water environments (> 25 °C).

Resorption of eggs	Ability to resorb all or part of the gametes instead of spawning them normally.
<u>Modalities:</u>	
<i>yes</i>	Organisms that have the ability of egg resorption.
<i>no</i>	Organisms that do not have the ability of egg resorption.
Sexual metamorphosis	Conspicuous change in the organism's body structure prior to reproduction.
<u>Modalities:</u>	
<i>yes</i>	Organisms that undergo sexual metamorphosis.
<i>no</i>	Organisms that do not undergo sexual metamorphosis.
Spawning frequency of the population	Period and frequency of spawning in a population.
<u>Modalities:</u>	
<i>continuous or semi-continuous</i>	Reproduction occurs all year round or for the most part of the year.
<i>annually; seasonal</i>	Yearly over a drawn out period of several weeks or a few months, or always in a defined season, peaks or epidemic swarming can occur within this period.
<i>multiple events/year</i>	More than once per year, but in relatively defined peaks or intense periods that do not fall within a drawn-out period.
Sperm type	Different types of sperm that occur in organisms and fertilize the eggs. Rouse and Jamieson (1987) proposed a system of classifying polychaete sperm based purely on function, using the terminology ect-aquasperm, ent-aquasperm and introsperm.
<u>Modalities:</u>	
<i>ect - aquasperm</i>	Type of sperm that are released into the water and fertilize similarly released eggs (Rouse 2005).
<i>ent - aquasperm</i>	Type of sperm that are released freely into the ambient water but differ from ect-aquasperm in being gathered by, or in some other way reaching, the female (Rouse 2005).
<i>introsperm</i>	Have no contact with water when passed from male to female (Rouse 2005).
Synchronization of spawning	Level of synchronization of the reproductive activity in a population.
<u>Modalities:</u>	
<i>yes</i>	Organisms whose populations undergo through a synchronized ripening of the gametes, usually culminating in an epidemic spawning event.
<i>no</i>	Organisms whose populations do not undergo through a synchronized ripening of the gametes.

Data coverage of traits

The amount of data available for the different traits varies greatly. The trait with the most entries is "Feeding structure", a morphological trait which is almost completely consistent within each family and known for all polychaete families. Information related to environmental preferences as well as mobility and feeding is likewise available for a large number of taxa, whereas many reproductive and larval traits are unknown for the majority

of polychaetes (Fig. 7). Most of the information is furthermore accompanied by a quotation of the original text passage which helps interpreting the assignment of the trait to the taxon and provides a means of quality control Fig. 7. A certain percentage of the missing text excerpts is however attributed to the fact that for a trait assigned according to an expert's judgement, no text excerpt can be recorded.

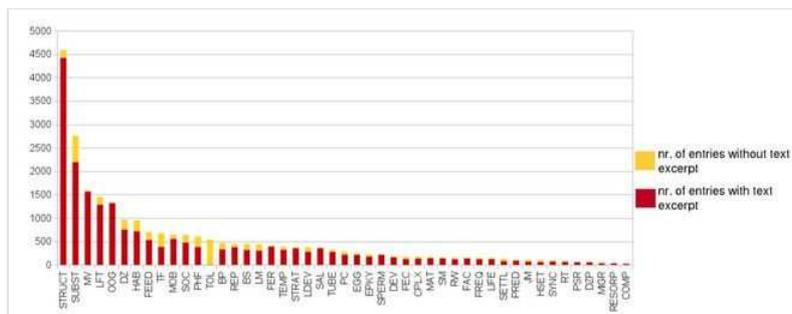


Figure 7.

Traits coverage with the total number of entries and the number of entries with and without text excerpt. Graph based on data in Suppl. material 9. Abbreviations of traits as follows: **BP** = Parental care / Brood protection; **BS** = Body size (max); **COMP** = Intra- and Interspecies competition; **CPLX** = Complex species; **DEV** = Developmental mechanism; **DZ** = Depth zonation (benthos); **DZP** = Depth zonation (pelagic); **EGG** = Egg size; **EPKY** = Epitoky; **FAC** = Factors triggering reproduction; **FEC** = Fecundity; **FEED** = Feeding type; **FER** = Fertilization type; **FREQ** = Spawning frequency of a population; **HAB** = Habitat type; **HSET** = Habitat type of settlement/early development; **JMOB** = Juvenile mobility; **LDEV** = Larval development; **LFT** = Larval feeding type; **LIFE** = Lifespan; **LM** = Larval mode of development; **MAT** = Age at first reproduction; **MIGR** = Migrations of adult; **MOB** = Mobility of adult; **MV** = Metamorphosis volume; **OOG** = Pattern of oogenesis; **PC** = Location of parental care; **PHF** = Physiographic feature; **PRED** = Predated by; **PSR** = Population sex ratio; **REP** = Mode of reproduction; **RESORP** = Resorption of eggs; **RT** = Reproduction temperature; **RW** = Ecosystem engineering; **SAL** = Survival salinity; **SETTL** = Substrate type of settlement; **SM** = Sexual metamorphosis; **SOC** = Sociability; **SPERM** = Sperm type; **STRAT** = Reproduction strategy of the individual; **STRUCT** = Feeding structure; **SUBST** = Substrate type; **SYNC** = Synchronization of spawning; **TEMP** = Survival temperature; **TF** = Typically feeds on; **TOL** = Tolerance (AMBI index); **TUBE** = Tube / burrow material.

Fig. 8 provides a more detailed overview of the amount of information available per taxonomic rank for each trait. For several traits, information is mainly available for taxonomic ranks higher than species (e.g. information about depth distribution is known for most families, but less so for individual species).

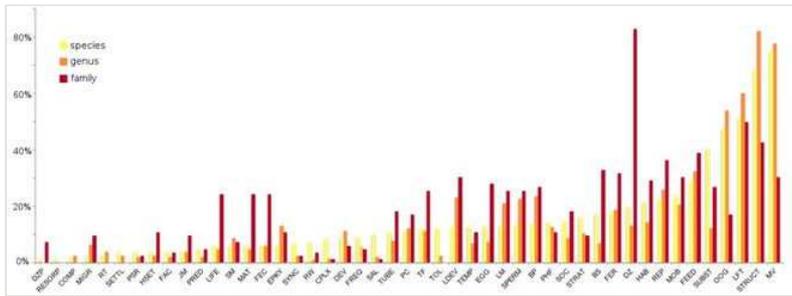


Figure 8.

Percentage of number of species, genus and families for which information is available, per trait. Graph based on data in Suppl. material 10. Abbreviations of traits as in Fig. 7.

Temporal coverage

Notes: To collect information on biological traits of polychaetes, literature from the past two centuries has been employed. However, the majority of the information was collected from literature published during the last decades Fig. 9.

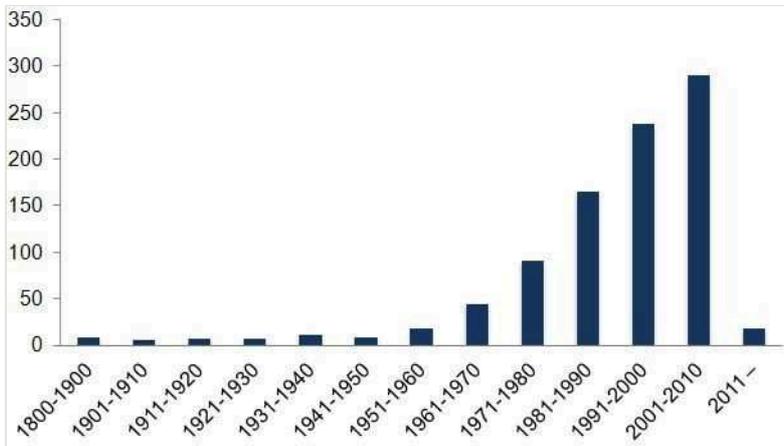


Figure 9.

Temporal distribution (publication year) of references used to collect polychaete trait information. URLs and other references without a publication year are excluded from the graph. Graph based on data in Suppl. material 11.

Usage rights

Use license: Open Data Commons Attribution License

IP rights notes: All data in the database can be freely used. Please cite this publication or the resource when using a large part of the data in your analyses. If individual records from the *Polytraits* database are displayed on a third-party website, it is recommended to cite also the record-level creator (see <http://polytraits.lifewatchgreece.eu/download> for more information), simply because often this information helps in assessing the quality of an entry (e.g. when "Expert's judgement" is given as a source for the information).

Data resources

Data package title: Polytraits database

Resource link: <http://polytraits.lifewatchgreece.eu>

Number of data sets: 1

Data set name: Polytraits database

Character set: UTF-8

Download URL: <http://polytraits.lifewatchgreece.eu/download>

Data format: Darwin Core, csv, MySQL

Description: The data in the *Polytraits* database can be accessed in five different ways:

1. Browsing the data on a taxon-by-taxon basis through the [Polychaetes Scratchpads](#) (tab "Polychaetes" – tab "Traits").
2. Browsing the data through the [Encyclopedia of Life](#) (EOL).
3. Downloading the data as a comma separated value (*.csv) file from the [Polytraits website](#). The download can be customised by submitting a list of taxa for which trait information should be retrieved or by limiting the output to specific traits. As an output format, either a *Polytraits*-specific format or DarwinCore can be chosen. Note that due to the restrictions of DarwinCore for trait data no information on the absence of a modality in a taxon could be included, this file therefore contains fewer records than the native export. A description of the returned fields for the *Polytraits*-specific format are documented in Table 4, those for the DarwinCore format are described below.
4. Downloading the full database as a [MySQL script](#), automatically created at monthly intervals. A full documentation of all database tables and fields, including an Entity Relationship Diagram is provided in Suppl. material 1. The script can be imported into any local MySQL database and will automatically create all relevant tables and their data.
5. Accessing the data programmatically via a web service ([REST API](#), application programming interface). This approach addresses a more technically oriented

audience. The API can be accessed as follows: **http://polytraits.lifewatchgreece.eu/{method}/{query}/{format}/?{other parameter key value pairs}**. Two methods are currently available (documented in Table 5 and Table 6; an exemplary client implementing both methods is provided both as a PHP and a perl script in Suppl. material 2):

1. **taxon**: searches for a taxon name and returns taxon identifier(s);
2. **traits**: retrieves traits information for one or more taxon identifiers.

Table 4.

List of fields returned by the *Polytraits* database, either when downloading information from the [Polytraits download page](#) (export format "default CSV file") or when accessing the database through the REST API.

Column label	Column description
Taxon	The taxon for which the information was recorded.
Author	The author and year of the <i>Taxon</i> for which the information was recorded.
Valid taxon	Currently accepted name of the <i>Taxon</i> (as stored in the <i>Polytraits</i> database – information might not be up to date with the latest taxonomic literature in some cases). If <i>Taxon</i> is currently accepted, this field contains the same value as <i>Taxon</i> .
Valid author	Currently accepted name of the <i>Author</i> (as stored in the <i>Polytraits</i> database – information might not be up to date with the latest taxonomic literature in some cases). If <i>Taxon</i> is currently accepted, this field contains the same value as <i>Author</i> .
Taxonomic status	Information on why <i>Taxon</i> is not currently valid (e.g. objective synonym). If <i>Taxon</i> and <i>Author</i> are currently accepted, this field is empty.
Source of synonymy	Literature reference for the <i>Taxonomic status</i> (if present).
Parent taxon	The <i>Taxon</i> 's direct parent in the taxonomic classification (as stored in the <i>Polytraits</i> database).
Trait	The biological trait for which information is available (e.g. "Feeding type").
Modality	The sub-category of the <i>Trait</i> for which information is available (e.g. "Carnivore").
Modality abbreviation	An abbreviated version of the often verbose <i>Modality</i> – useful as a label in further analyses of the data (e.g. "FEED_C").
Traitvalue	Describes the affinity of the <i>Taxon</i> to the above <i>Modality</i> – can be either "0" (for "absent") or "1" (for "present").
Reference	Literature reference leading to the assignment of the <i>Traitvalue</i> to the <i>Modality</i> for the <i>Taxon</i> .
DOI	Digital Object Identifier (where available) of the <i>Reference</i> .
Value creator	Person who assigned the <i>Traitvalue</i> to the <i>Modality</i> for the <i>Taxon</i> , supported by a <i>Reference</i> .
Value creation date	Date and time when the above information was entered into the database.
Text Excerpt	A quotation of the original text passage from the literature source that led to the assignment of assignment of the <i>Modality</i> / <i>Traitvalue</i> to the <i>Taxon</i> . Empty if information has not been recorded yet.
Text Excerpt creator	Person who entered the <i>Text Excerpt</i> . Only present if <i>Text Excerpt</i> is present.
Text Excerpt creation date	Date and time when the <i>Text Excerpt</i> was entered into the database. Only present if <i>Text Excerpt</i> is present.

Table 5.

A method to search for taxa in the *Polytraits* database. Returns matching taxon identifiers.

URL: <http://polytraits.lifewatchgreece.eu/taxon/{query}/{format}?{other parameter key value pairs}>

Example: Retrieve all taxon information about taxa starting with "Syllis" and return them as a numeric array: <http://polytraits.lifewatchgreece.eu/taxon/Syllis/json/?exact=0&verbose=1&assoc=0>

Parameter name	data type	allowed values	default value	Parameter Note	Output
query	string	any taxon name		spaces have to be url-encoded, case-insensitive	
format	string	json xml	json	Optional. Currently only json is implemented	
exact	boolean	1 0	1	optional	When true, returns only the exact match for the query string. When false, returns all matches beginning with the query string.
verbose	boolean	1 0	0	optional	When true, returns per taxon: <i>array (taxon ID, taxon name, authority, valid taxon ID, valid taxon name, valid authority, status, source of synonymy, rank)</i> . When false, returns per taxon: <i>array (taxon ID, taxon name)</i> .
assoc	boolean	1 0	0	optional	When true, an associative array is returned with the taxon ID as keys. When false, a numeric array is returned.

Table 6.

A method to retrieve trait information for one or more taxon identifiers from the *Polytraits* database. Returns all trait information for the given IDs.

URL: `http://polytraits.lifewatchgreece.eu/traits/{query}/{format}?{other parameter key value pairs}`

Example: Retrieve all trait information about the taxa with ID 1 and 2 and return them as an associative array with the taxon IDs as keys: <http://polytraits.lifewatchgreece.eu/traits/1,2/json/?verbose=1&assoc=1>

The output fields are documented in Table 4.

Parameter name	data type	allowed values	default value	Parameter Note	Output
query	integer or string	one or more taxon identifiers, comma-separated		Use "taxon" method to retrieve a list of IDs. Up to 10 Ids are allowed	
format	string	json xml	json	Optional. Currently only json is implemented	
verbose	boolean	1 0	1	Optional	When true, returns per taxon the fields documented in Table 4. When false, returns per taxon: <i>array (trait, modality, traitvalue)</i> .
assoc	boolean	1 0	1	optional	When true, an associative array is returned with the taxon ID as keys. When false, a numeric array is returned.

Whereas data downloaded as csv files or accessed through the API always reflect the latest changes in the database, the MySQL export is provided as monthly snapshots. The data available through EOL are a one-time export and reflect the database contents as of November 6th, 2013.

Column label	Column description
scientificName	The taxon for which the information was recorded.
scientificNameAuthorship	The author and year of the taxon for which the information was recorded.
acceptedNameUsage	Currently accepted name and authorship of the scientificName (as stored in the Polytraits database – information might not be up to date with the latest taxonomic literature in some cases).
taxonomicStatus	The status of the use of the scientificName (e.g. objective synonym, subjective synonym) as stored in the Polytraits database. Empty if scientificName is the currently accepted name.
MeasurementOrFact	Unique Resource Identifier pointing to the definition of a biological trait.

measurementValue	Unique Resource Identifier pointing to the definition of a modality (trait category) which is expressed by the current taxon.
dcterms:bibliographicCitation	Full literature reference (including Digital Object Identifier (DOI) where present) supporting the trait information for the current taxon.
measurementRemarks	A quotation of the original text passage containing the trait information for the current taxon.
measurementDeterminedBy	Person who entered the trait information for this taxon into the database.
measurementDeterminedDate	Date the trait information was entered into the database or last modified.

Additional information

Resource citation: Polytraits Team (2013 –) *Polytraits*: A database on biological traits of polychaetes, 25034 data records, Contributors: Faulwetter S, Markantonatou V, Pavlouidi C, Papageorgiou N, Keklikoglou K, Chatzinikolaou E, Pafilis E, Chatzigeorgiou G, Vasileiadou K, Dailianis T, Fanini L, Koulouri P, Arvanitidis C. Online at <http://polytraits.lifewatchgreece.eu>. Version 1.0, last updated on: 2013-11-14.

Acknowledgements

The authors thank Lyubomir Penev (Pensoft Publishers), Cynthia Parr and Jennifer Hammock (Encyclopedia of Life) for discussions on publishing and disseminating trait data and for publishing the *Polytraits* data through the Encyclopedia of Life. Simon Rycroft (NHM London) is thanked for support during the development of the Scratchpads module which allows the display of trait information in Scratchpads.

References

- Almeida WD, Christoffersen ML, Amorim DD, Garaffoni AR, Silva GS (2003) Polychaeta, Annelida and Articulata are not monophyletic: articulating the Metameria (Metazoa, Coelomata). *Revista Brasileira de Zoologia* 20: 23-57. DOI: [10.1590/S0101-81752003000100006](https://doi.org/10.1590/S0101-81752003000100006)
- Arvanitidis C (1994) Συστηματική και βιονομική μελέτη των μακροβενθικών πολυχαίτων (Annelida, Polychaeta) του βορείου Αιγαίου. [Systematic and bionomic study of the macrobenthic Polychaetes (Annelida) on the northern Aegean]. Aristotelian University of Thessaloniki, Thessaloniki, 512 pp. [In Greek]. URL: <http://phdtheses.ekt.gr/eadd/handle/10442/5045>
- Ballesteros E (2006) Mediterranean coralligenous assemblages: a synthesis of present knowledge. *Mediterranean coralligenous assemblages: a synthesis of present knowledge* 44: 123-196.
- Bevilacqua S, Terlizzi A, Claudet J, Fraschetti S, Boero F (2012) Taxonomic relatedness does not matter for species surrogacy in the assessment of community responses to environmental drivers. *Journal of Applied Ecology* 49 (2): 357-366. DOI: [10.1111/j.1365-2664.2011.02096.x](https://doi.org/10.1111/j.1365-2664.2011.02096.x)

- Birkett D, Maggs C, Dring M (1998) Maerl (volume V). An overview of dynamic and sensitivity characteristics for conservation management of marine SACs. Scottish Association for Marine Science (UK Marine SACs Project), 116 pp.
- Bishop CD, Erezyilmaz DF, Flatt T, Georgiou CD, Hadfield MG, Heyland A, Hodin J, Jacobs MW, Maslakova SA, Pires A, Reitzel AM, Santagata S, Tanaka K, Youson JH (2006) What is metamorphosis? *Integrative and Comparative Biology* 46 (6): 655-661. DOI: [10.1093/icb/icl004](https://doi.org/10.1093/icb/icl004)
- Blake J, Arnofsky P (1999) Reproduction and Larval Development of the Spioniform Polychaeta with Application to Systematics and Phylogeny. *Hydrobiologia* 402: 57-106. DOI: [10.1023/A:1003784324125](https://doi.org/10.1023/A:1003784324125)
- Borja A, Franco J, Pérez V (2000) A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments. *Marine Pollution Bulletin* 40 (12): 1100-1114. DOI: [10.1016/S0025-326X\(00\)00061-8](https://doi.org/10.1016/S0025-326X(00)00061-8)
- Bremner J (2008) Species' traits and ecological functioning in marine conservation and management. *Journal of Experimental Marine Biology and Ecology* 366: 37-47. DOI: [10.1016/j.jembe.2008.07.007](https://doi.org/10.1016/j.jembe.2008.07.007)
- Bremner J, Rogers S, Frid C (2003) Assessing functional diversity in marine benthic ecosystems: a comparison of approaches. *Marine Ecology Progress Series* 254: 11-25. DOI: [10.3354/meps254011](https://doi.org/10.3354/meps254011)
- Carson HS, Hentschel BT (2006) Estimating the dispersal potential of polychaete species in the Southern California Bight: implications for designing marine reserves. *Marine Ecology Progress Series* 316: 105-133. DOI: [10.3354/meps316105](https://doi.org/10.3354/meps316105)
- Cooper K, Barrio Froján CS, Defew E, Curtis M, Fleddum A, Brooks L, Paterson D (2008) Assessment of ecosystem function following marine aggregate dredging. *Journal of Experimental Marine Biology and Ecology* 366 (1): 82-91. DOI: [10.1016/j.jembe.2008.07.011](https://doi.org/10.1016/j.jembe.2008.07.011)
- Dorresteyn AW, Westheide W (1999) Reproductive strategies and developmental patterns in annelids. 142. Springer, 328 pp. DOI: [10.1007/978-94-017-2887-4](https://doi.org/10.1007/978-94-017-2887-4)
- Eckelbarger K (1983) Evolutionary radiation in polychaete ovaries and vitellogenic mechanisms: their possible role in life history patterns. *Canadian Journal of Zoology* 61 (3): 487-504. DOI: [10.1139/z83-065](https://doi.org/10.1139/z83-065)
- Eckelbarger KJ (2005) Oogenesis and Oocytes. *Hydrobiologia* 535/536: 179-198. DOI: [10.1007/s10750-004-4397-y](https://doi.org/10.1007/s10750-004-4397-y)
- Eleftheriou M (1997) *Aqualex: a glossary of aquaculture terms*. Wiley-Praxis Series in Aquaculture and Fisheries, 410 pp.
- Fauchald K, Jumars P (1979) The diet of worms: a study of polychaete feeding guilds. *Oceanography and Marine Biology: An Annual Review* 17: 193-284.
- Fauvel P (1927) *Faune de France: Polychètes sédentaires*. Lechevalier, Paris, 494 pp. [In French]. URL: [http://www.faunedefrance.org/bibliotheque/docs/P.FAUVEL\(FdeFr16\)Polychetes-sedentaires.pdf](http://www.faunedefrance.org/bibliotheque/docs/P.FAUVEL(FdeFr16)Polychetes-sedentaires.pdf)
- George JD (1966) Reproduction and early development of the spionid polychaete, *Scolecoides viridis* (Verrill). *The Biological Bulletin* 130 (1): 76-93. DOI: [10.2307/1539954](https://doi.org/10.2307/1539954)

- Hall B, Olson W (2003) *Keywords and Concepts in Evolutionary Developmental Biology*. Harvard University Press, Cambridge, 496 pp.
- Hartmann-Schröder G (1996) *Annelida, Borstenwürmer, Polychaeta*. Gustav Fischer Verlag, Jena, 648 pp. [In German].
- Holthe T (1986) *Polychaeta Terebellomorpha*. *Marine Invertebrates of Scandinavia* 7: 1-194.
- Holt T, Rees E, Hawkins S, Seed R (1998) *Biogenic Reefs (volume IX)*. An overview of dynamic and sensitivity characteristics for conservation management of marine SACs. Scottish Association for Marine Science (UK Marine SACs Project), 170 pp.
- Jones C, Lawton J, Shachak M (1994) Organisms as ecosystem engineers. *Oikos* 69: 373-386. DOI: [10.2307/3545850](https://doi.org/10.2307/3545850)
- Klein K, Santamarina JC (2005) Soft Sediments: Wave-based Characterization. *International Journal of Geomechanics* 5: 147-157. DOI: [10.1061/\(ASCE\)1532-3641\(2005\)5:2\(147\)](https://doi.org/10.1061/(ASCE)1532-3641(2005)5:2(147))
- Kristensen E, Penha-Lopes G, Delefosse M, Valdemarsen T, Quintana C, Banta G (2012) What is bioturbation? The need for a precise definition for fauna in aquatic sciences. *Marine Ecology Progress Series* 446: 285-302. DOI: [10.3354/meps09506](https://doi.org/10.3354/meps09506)
- Kupriyanova EK, Nishi E, ten Hove HA, Rzhavsky AV (2001) A review of life history in serpulimorph polychaetes: ecological and evolutionary perspectives. *Oceanography and Marine Biology: an Annual Review* 39: 1-101.
- Levin LA, Bridges TS (1995) Pattern and diversity in reproduction and development. In: McEdward L (Ed.) *Ecology of Marine Invertebrate Larvae*. CRC Press, Boca Raton, New York, London, Tokyo, 48 pp.
- Lincoln R, Boxshall G, Clark P (1998) *A dictionary of ecology, evolution and systematics*. 2nd. Cambridge University Press, Cambridge, 361 pp.
- MarLIN (2006) BIOTIC - Biological Traits Information Catalogue. Cited: 08/11/2013. Marine Life Information Network. Plymouth: Marine Biological Association of the United Kingdom. URL: www.marlin.ac.uk/biotic
- MarLIN (2013) Marine Life Information Network. cited:05/11/2013. Plymouth: Marine Biological Association of the United Kingdom. URL: www.marlin.ac.uk
- Mayr E, Ashlock PD (1991) *Principles of Systematic Zoology*. 2nd. McGraw-Hill, New York, 475 pp.
- McEdward L (1995) *Ecology of Marine Invertebrate Larvae*. CRC Press, Boca Raton, New York, London, Tokyo, 464 pp.
- McHugh D, Fong P (2002) Do life history traits account for diversity of polychaete annelids? *Invertebrate Biology* 121 (4): 325-338. DOI: [10.1111/j.1744-7410.2002.tb00133.x](https://doi.org/10.1111/j.1744-7410.2002.tb00133.x)
- Mouillot D, Mason WHN, Dumay O, Wilson JB (2005) Functional regularity: a neglected aspect of functional diversity. *Oecologia* 142 (3): 353-359. DOI: [10.1007/s00442-004-1744-7](https://doi.org/10.1007/s00442-004-1744-7)
- Nielsen C (2000) The origin of metamorphosis. *Evolution and Development* 2 (3): 127-129. DOI: [10.1046/j.1525-142x.2000.00057.x](https://doi.org/10.1046/j.1525-142x.2000.00057.x)

- Nielsen C (2009) How Did Indirect Development With Planktotrophic Larvae Evolve? *Biological Bulletin* 216: 203-215. URL: <http://www.biolbull.org/content/216/3/203.short>
- Paganelli D, Marchini A, Occhipinti-Ambrogi A (2012) Functional structure of marine benthic assemblages using Biological Traits Analysis (BTA): A study along the Emilia-Romagna coastline (Italy, North-West Adriatic Sea). *Estuarine, Coastal and Shelf Science* 96: 245-256. DOI: [10.1016/j.ecss.2011.11.014](https://doi.org/10.1016/j.ecss.2011.11.014)
- Qian P, Dahms H (2006) Larval ecology of the Annelida. In: Jamieson BG (Ed.) *Reproductive Biology and Phylogeny of Annelida*. CRC Press, 698 pp.
- Read G (2006) *Annelid Resources*. URL: <http://www.annelida.net>
- Rouse G (2005) Annelid Sperm and Fertilization Biology. *Hydrobiologia* 535/536: 167-178. DOI: [10.1007/s10750-004-4390-5](https://doi.org/10.1007/s10750-004-4390-5)
- Rouse G, Jamieson BM (1987) An ultrastructural study of the spermatozoa of the polychaetes *Eurythoe complanata* (Amphinomidae), *Clymenella* sp. and *Micromaldane* sp. *Journal of Submicroscopical Cytology* 19: 573-584.
- Rouse GW, Pleijel F (2001) *Polychaetes*. Oxford University Press, Oxford, 354 pp.
- Rouse GW, Pleijel F (Ed.) (2006) *Reproductive Biology and Phylogeny of Annelida*. Science Publishers, New Hampshire, 688 pp.
- Ryan F (2011) *Metamorphosis: Unmasking the mystery of how life transforms*. Oneworld Publications, 320 pp.
- San Martín G (2003) *Annelida Polychaeta II: Syllidae*. Museo Nacional de Ciencias Naturales, Madrid, 554 pp. [In Spanish].
- Schleuter D, Daufresne M, Massol F, Argillier C (2010) A user's guide to functional diversity indices. *Ecological Monographs* 80 (3): 469-484. DOI: [10.1890/08-2225.1](https://doi.org/10.1890/08-2225.1)
- Smiseth P, Kölliker M, Royle N (2012) What is parental care? In: Royle NJ, Smiseth PT, Kölliker M (Eds) *The evolution of parental care*. Oxford University Press DOI: [10.1093/acprof:oso/9780199692576.003.0001](https://doi.org/10.1093/acprof:oso/9780199692576.003.0001)
- Smith V, Rycroft S, Harman K, Scott B, Roberts D (2009) Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life. *BMC Bioinformatics* 10: S6. DOI: [10.1186/1471-2105-10-S14-S6](https://doi.org/10.1186/1471-2105-10-S14-S6)
- Sunamura T (1992) *Geomorphology of rocky coasts (Coastal Morphology and Research Series)*. John Wiley & Sons, Chichester; New York, 302 pp.
- Tyler EM, Somerfield P, Berghe EV, Bremner J, Jackson E, Langmead O, Palomares M, Webb T (2012) Extensive gaps and biases in our knowledge of a well-known fauna: implications for integrating biological traits into macroecology. *Global Ecology and Biogeography* 21 (9): 922-934. DOI: [10.1111/j.1466-8238.2011.00726.x](https://doi.org/10.1111/j.1466-8238.2011.00726.x)
- Tzetlin A, Purschke G (2005) Pharynx and intestine. *Hydrobiologia* 535-536 (1): 199-225. DOI: [10.1007/s10750-004-1431-z](https://doi.org/10.1007/s10750-004-1431-z)
- Usseglio-Polatera P, Bournaud M, Richoux P, Tachet H (2000) Biomonitoring through biological traits of benthic macroinvertebrates: how to use species trait databases. *Hydrobiologia* 422/423: 153-162. DOI: [10.1007/978-94-011-4164-2_12](https://doi.org/10.1007/978-94-011-4164-2_12)
- Vieitez J, Alos C, Parapar J, Besteiro C, Moreira J, Núñez J, Laborda A, San Martín G (2004) *Fauna Iberica, Vol. 25. Annelida Polychaeta I*. Museo Nacional de Ciencias Naturales, Madrid, 530 pp. [In Spanish].

- Wilson WH (1991) Sexual reproductive modes in polychaetes: classification and diversity. *Bulletin of Marine Science* 48: 500-516. URL: <http://www.ingentaconnect.com/content/umrsmas/bullmar/1991/00000048/00000002/art00037>
- Wolff WJ (1973) The estuary as a habitat: an analysis of data on the soft-bottom macrofauna of the estuarine area of the Rivers Rhine, Meuse, and Scheldt. Rijksuniversiteit te Leiden, Leiden, 242 pp.

Supplementary materials

Suppl. material 1: Database documentation

Authors: Sarah Faulwetter

Data type: pdf document

Brief description: Document describing all database tables that are made available to the user for download through the *Polytraits* website. Contains both a detailed description of all database fields as well as a diagram of the tables' relationships.

Filename: polytraits_db_documentation_2013-11-05.pdf - [Download file](#) (945.44 kb)

Suppl. material 2: Polytraits REST API

Authors: Sarah Faulwetter

Data type: php / perl script (zipped)

Brief description: Exemplary PHP and perl client demonstrating the two data retrieval methods currently implemented by the *Polytraits* database.

Filename: polytraits_client.php.zip - [Download file](#) (2.34 kb)

Suppl. material 3: Full list of literature references

Authors: Sarah Faulwetter

Data type: csv file with literature references

Brief description: Full list of all literature references from which data for the *Polytraits* database were collected.

Filename: polytraitsReferences13-10-15.csv - [Download file](#) (468.95 kb)

Suppl. material 4: Literature sources used.

Authors: Sarah Faulwetter

Data type: Excel file with summarized data from database

Brief description: Types of literature sources used in the *Polytraits* database.

Filename: figure_2a.xls - [Download file](#) (6.00 kb)

Suppl. material 5: Origin of trait information

Authors: Sarah Faulwetter

Data type: Sarah Faulwetter

Brief description: Origin (literature type) of the traits information in the database.

Filename: figure_2b.xls - [Download file](#) (6.00 kb)

Suppl. material 6: Taxon ranks and status

Authors: Sarah Faulwetter

Data type: Excel file with summarized data from database

Brief description: Taxonomic rank and status of taxa in the *Polytraits* database.

Filename: figure_4.xls - [Download file](#) (7.00 kb)

Suppl. material 7: Number of species

Authors: Sarah Faulwetter

Data type: Excel file with summarized data from database

Brief description: Number of valid species contained in the *Polytraits* database, per family.

Filename: figure_5.xls - [Download file](#) (20.00 kb)

Suppl. material 8: Number of records

Authors: Sarah Faulwetter

Data type: Excel file with summarized data from database

Brief description: Number of taxon-modality-value records on family, genus and species level, per family.

Filename: figure_6.xls - [Download file](#) (18.00 kb)

Suppl. material 9: Traits coverage

Authors: Sarah Faulwetter

Data type: Excel file with summarized data from database

Brief description: Traits coverage with the total number of entries and the number of entries with and without text excerpts.

Filename: figure_7.xls - [Download file](#) (12.00 kb)

Suppl. material 10: Completeness of traits

Authors: Sarah Faulwetter

Data type: Excel file with summarized data from database

Brief description: Percentage of number of species, genus and families for which information is available, per trait.

Filename: figure_8.xls - [Download file](#) (15.50 kb)

Suppl. material 11: Temporal distribution of literature

Authors: Sarah Faulwetter

Data type: Excel file with summarized data from database

Brief description: Temporal distribution (publication year) of references which were used in the *Polytraits* database.

Filename: Figure_9.xls - [Download file](#) (10.50 kb)

Chapter 5

Faulwetter S, Papageorgiou N, Koulouri Y, Fanini L,
Chatzinikolaou E, Markantonatou V, Pavludi C,
Chatzigeorgiou G, Keklikoglou K, Vasileiadou K, Basset A,
Pinna M, Rosati I, Reizopoulou S, Nicolaidou A, Arvanitidis
C. Comparison of structural and functional stability of
polychaete assemblages in coastal lagoons. *Journal of Sea
Research*, submitted.

Comparison of structural and functional stability of polychaete assemblages in coastal lagoons

Introduction

The increasing global loss of biodiversity is a primary concern not only for ecologists, but for society at large who profits from the goods and services that ecosystems provide. Although it is considered that the loss of species has negative effects on ecosystem functioning, the biodiversity-ecosystem function (BEF) relationship still is the subject of an ongoing debate (Boero and Bonsdorff 2007, Hooper et al. 2005, Lawton and Brown 1993, Naeem and Wright 2003, Naeem et al. 2002 2009, O'Connor and Crowe 2005). Central questions are, for example, whether high species diversity provides stability of ecosystems, how the loss of biodiversity alters ecosystem functioning and how many species extinctions an ecosystem can tolerate before it becomes “unstable” (Ives and Carpenter 2007, McCann 2000). However, the concept of “stability” has received a wide range of definitions (Loreau et al. 2001, McCann 2000) and the relationship between diversity and function is complex and still poorly understood (Larsen et al. 2005, Mayfield et al. 2010, McCann 2000, Naeem and Wright 2003). Over time, different hypotheses have been postulated to explain this relationship. Among these are a) the redundancy hypothesis, postulating that species overlap in their function to a certain degree and can thus compensate for species loss (Walker 1992), b) the rivet hypothesis which claims that all species are unique and that their loss will have consequences for the functioning of an ecosystem (Ehrlich and Ehrlich 1981), and c) the idiosyncratic hypothesis, which states that ecosystem processes change when diversity changes, but changes are unpredictable and context-dependent (Naeem et al. 1995 2002). Since it is assumed that species loss results in functional loss, many studies have used structural biodiversity patterns as a surrogate of ecosystem function (see Mayfield et al. 2010, Srivastava and Vellend 2005 for an overview), but this relationship is not linear and seems to be determined by a number of factors which are context-dependent and not fully understood (Symstad et al. 2003, Tilman et al. 1997a).

In the terrestrial realm, much research has been dedicated to the effect of species loss on ecosystem functioning, either through experimentally altered diversity (e.g. Díaz et al. 2003, Hooper and Vitousek 1997, Tilman et al. 1997a) or through theoretical approaches simulating extinction scenarios (e.g. Fonseca and Ganade 2001, Petchey and Gaston 2002, Petchey et al. 2004). In the marine realm, however, the relationship between biodiversity and ecosystem

functioning has only recently started to appear on the research agenda (Boero and Bonsdorff 2007, Covich et al. 2004, Danovaro and Pusceddu 2007, Gamfeldt and Hillebrand 2008, Naeem 2006, O'Connor and Crowe 2005, Solan et al. 2006), and few studies have compared structural and functional patterns (e.g. Bremner 2008, Bremner et al. 2003, 2006, Hewitt et al. 2008, Marchini et al. 2008, Solan et al. 2004, Törnroos and Bonsdorff 2012). The ability to predict effects of species loss and to understand which intrinsic factors are responsible for stabilising communities is, however, of utmost importance to conserve ecosystem functioning (O'Connor and Crowe 2005, Petchey 2000). Towards this end, Clarke and Warwick (1998) proposed a method to quantify structural redundancy in marine benthic species assemblages by identifying small subsets of species in a community that revealed the same multivariate patterns of biodiversity as the complete set of species. From a structural point of view, the species not included in this subset can be considered redundant, in the sense that they are not required to reproduce the community patterns derived from the full dataset. However, the authors explicitly state that the species would have also to be functionally interchangeable in order to use the method as a tool for assessing the resilience or compensation potential of a community (Clarke and Warwick 1998).

In an attempt to shed light on the relationship between multivariate patterns of community structure and functioning, the present study adopts the approach of Clarke and Warwick (1998) and combines it with a Biological Traits Analysis (BTA). The BTA is an approach that uses species' biological traits as indicators of functioning in order to cover multiple aspects of the interactions and functions performed by the biological components of an ecosystem (Bremner et al. 2003). By applying Clarke and Warwick's (1998) approach to the biological traits data instead of the community structure data, *functionally* redundant species can be identified. In this study, the above method is then taken a step further to explore the changes of both structural and functional multivariate patterns under simulations of random species loss. The resulting patterns indicate the magnitude of structural and functional changes that might follow species loss and provide information about the potential stability of the ecosystem, since any real changes are likely to fall into the limits of these simulations.

The concept is tested by assessing the structural and functional patterns of polychaete communities in six coastal lagoons in the Mediterranean and Black Sea. Coastal lagoons are functional ecotone ecosystems (Basset et al. 2013) with strong seasonal changes, often subjected to intensive anthropogenic use and pollution and eutrophication (e.g. Arvanitidis et al. 2005,

Barnes 1980, 1991). They often are characterised by reduced species diversity (e.g. Mouillot et al. 2007, Reizopoulou and Nicolaidou 2004), making them an intriguing environment to study ecosystem processes. The lagoons studied here contain varying numbers of species (from 5 to 21) which allows assessing the importance of species richness versus species identity for the stability of structural and functional patterns. Systems with low species richness are furthermore considered ideal for examining the biodiversity-ecosystem functioning relationship since the roles of individual species can be studied (Vitousek and Hooper 1993).

The analyses in this study are performed under the assumption that a) the species pool of each lagoon is not altered by re-entry of species from the adjacent sea upon loss of species from the lagoon and b) that species interactions do not change upon removal of certain species, and abundant and rare species continue as such after random removal of other species. Furthermore, only the polychaetes fraction of the macrobenthos dataset was analysed. Polychaetes are a dominant taxon in macrobenthic assemblages in terms of abundance, biomass and species richness. They play an important role in the stability and functioning of benthic communities (Fauchald 1977, Hutchings 1998) and have often be used as surrogates for the total benthic fauna in environmental studies (Giangrande et al. 2005, Hutchings 1998). Polychaetes also express a very large diversity of functional traits, and their ecology is very well studied. By focusing on a rather well-studied taxon with plenty of available expertise to collect traits information, the bias due to gaps in information – a common problem in biological traits analysis (Tyler et al. 2011) – is minimised.

The approach followed in this study does not attempt to determine absolute effects of the loss of biodiversity on ecosystem function, but it rather investigates how functional patterns in time and space might be affected by the loss of species, and whether these changes can be explained by any of the aforementioned hypotheses (i.e. redundancy, rivet or idiosyncratic hypothesis). Furthermore, through random species combinations it is possible to identify structural and functional “drivers” and “passengers” (Walker 1992) in the assemblage, i.e. those species that contribute most or least to the observed patterns. Finally, by comparing the structural and functional patterns, it is then possible to determine to what extent structural redundancy can be used as a proxy for functional redundancy.

Material and methods

Datasets

Species occurrence data / Sampling information / Environmental information

The biotic dataset contains occurrence data of 48 polychaete species from six coastal lagoons in the Eastern Mediterranean Sea: Grado Marano, Grado Valle di Pesca and Margherita di Savoia in Italy, Logarou and Agiasma in Greece, Varna in Bulgaria (Figure 1, Table 1). Taxonomic lists of benthic macroinvertebrates from the studied lagoons are available elsewhere (Barbone et al. 2007, Basset et al. 2012). All lagoons are characterised by natural seasonal fluctuations of salinity, temperature and oxygen concentration, but the degree of these fluctuations as well as the nature and level of additional anthropogenic pressures differ in each of these lagoons. The lagoons of Grado Marano, Agiasma and Varna are exposed to high organic and nutrient loading as well as chemical pollution and physical disturbances through navigation (Basset et al. 2008). Grado Valle di Pesca and Logarou are, at least in parts, used for fish farming. They are subjected, at least in parts, to increased organic and nutrient loading. Margherita di Savoia is used for salt production and receives strong fluctuations of salinity (from 40 to over 80 psu) over the course of the year, depending on the control of seawater influx and evaporation. The surface area of the six lagoons ranges from >100km (Grado Marano) to <1km (Grado Valle di Pesca). Details on lagoonal characteristics and environmental parameters can be found in the literature (Barbone et al. 2007, Basset et al. 2008, Reizopoulou et al. 2013, Sangiorgio et al. 2008, Sigala et al. 2012).

In each lagoon, four to five stations or station pairs were sampled along the most pronounced environmental gradient (salinity or pollution). Sampling took place once in autumn 2004 and once in spring 2005 (with the exception of Varna, for which only data for five stations from autumn are available). At each station and in each season, five replicate units were collected by means of a box corer. Polychaete abundances were subsequently averaged for the purposes of this analysis. Details on the sampling procedure and laboratory analyses can be found in Sigala et al. (2012).

Trait data

Biological Trait Analysis is based on the selection of indicator traits which serve as a proxy for the functional characteristics of the community and to which species are then assigned (Bremner 2008). For the purposes of this study, 31 biological traits were included, describing

the environmental preferences as well as morphological, behavioural and life cycle characteristics of each species. By encompassing a wide spectrum of aspects, a fairly accurate picture of ecological functioning of the polychaete assemblages can be obtained. Each of the traits is divided into several categories (e.g. the trait “mobility of adult” contains the categories “crawler”, “swimmer”, “burrower” and “sessile/semi-motile”). In total, 118 trait categories were included in the dataset (Supplementary materials 2, 3). Once the traits were selected, information for each species in the dataset was collected from the literature. A dedicated database with a web-based interface had been designed to facilitate data entry and manage the literature references (see Faulwetter et al. 2014 for a detailed description of the data and definition of the traits). Information on the expression of a trait in a species was entered by coding each category within that trait as either “present” or “absent” (or both, if the respective information was found in the literature). In total, 546 literature references were used to collect the information and experts' knowledge was employed to complement these data.

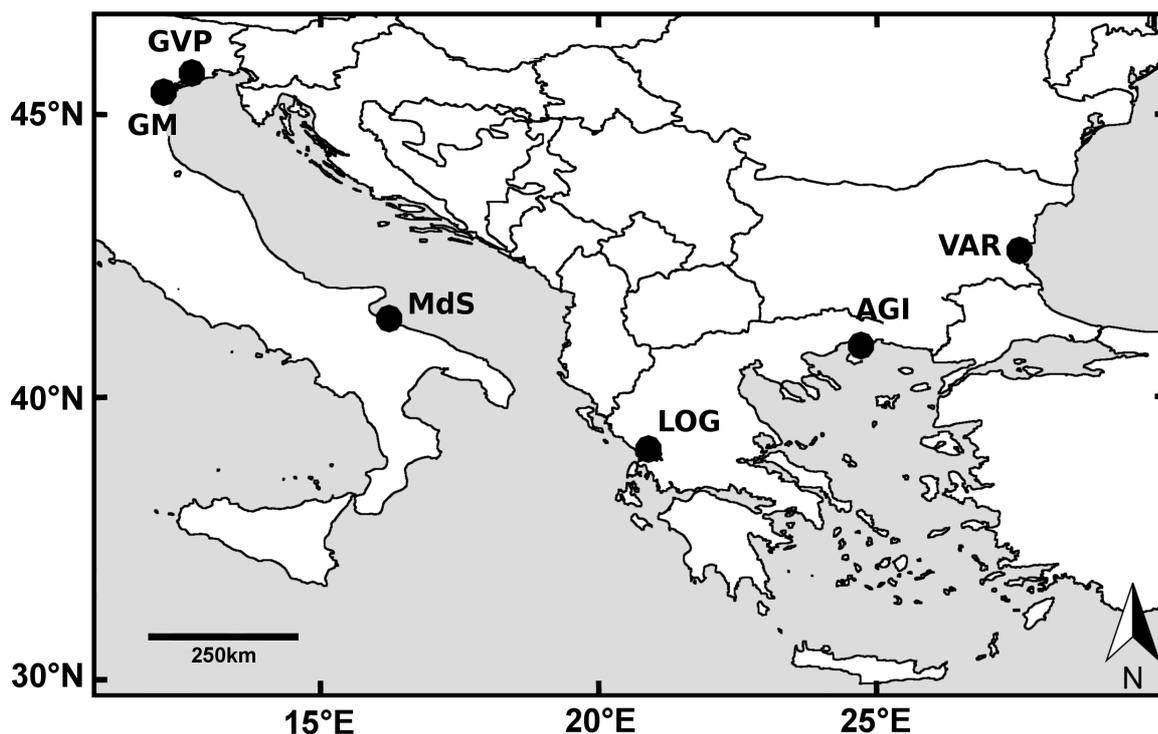


Figure 1: Map of the studied lagoons. Abbreviations: GM = Grado Marano, GVP = Grado Valle di Pesca, MdS = Margherita di Savoia, LOG = Logarou, AGI = Agiasma, VAR = Varna

Knowledge on the biological traits of species is incomplete for most marine invertebrates (Tyler et al. 2011), so even after the exhaustive literature search information was missing for several

traits in some species. In these cases (15.5% of the data), information was stepwise first deduced from congeneric species, then from confamiliar species. Since the taxonomy of polychaetes is often closely related to functional attributes (e.g. in certain families all members are sessile or follow a certain reproductive strategy), the approach was considered to be fairly robust towards introducing incorrect values. Naturally, a certain amount of (unknown) errors, especially in the more diverse families (e.g. the family Syllidae exhibits an extreme variety of reproductive modes) cannot be avoided, however, this strategy was still considered preferable over alternative approaches such as introducing a different code for “unknown”. This strategy resulted in a species-by-traits matrix containing information for all species. The matrix was then combined with the species-by-sample matrix (“sample” is defined here as the average of five replicate units per station per season) to determine which traits were expressed by the species in each sample: For each trait category and each sample, the abundances of all species exhibiting that trait category were summed up, resulting in a traits-by-sample matrix which was then subjected to analysis.

Table 1: Characteristics of the studied lagoons. S = Species number, N = total number of individuals in all samples, FR= Functional richness (number of expressed trait categories)

Lagoon	Environmental characterisation	Nr. of samples taken	S	N	FR
Agiasma	chemical and organic pollution	8 (1 without species)	9	947	109
Grado Marano	chemical and organic pollution	16 (4 without species)	15	140	110
Grado Valle di Pesca	fish farming	16	12	106	106
Logarou	fish farming	16	14	614	106
Margherita di Savoia	salt pans, hypersaline	16 (6 without species)	5	173	95
Varna	chemical and organic pollution	5	22	2344	115

Analyses

Simulations of species loss

For each lagoon, both the species-by-sample and the traits-by-sample matrices were fourth-root transformed and a resemblance matrix was calculated using the Bray-Curtis coefficient (Bray and Curtis 1957). Subsequently, a scenario of random species loss was simulated. For the

species-by-sample matrix, one species was excluded from the initial dataset, a resemblance matrix was calculated from this new dataset as described above, and this resemblance matrix was compared to the resemblance matrix of the full set of species through a mantle test using Spearman's rank correlation coefficient ρ . Species richness was reduced stepwise until only a single species was left. At each of these species richness levels, 100 random exclusions were performed (Figure 2), at high and low species richness levels this results in several species combinations being drawn multiple times since the number of possible combinations was lower than 100. The coefficient ρ of each comparison was then plotted against the species richness level.

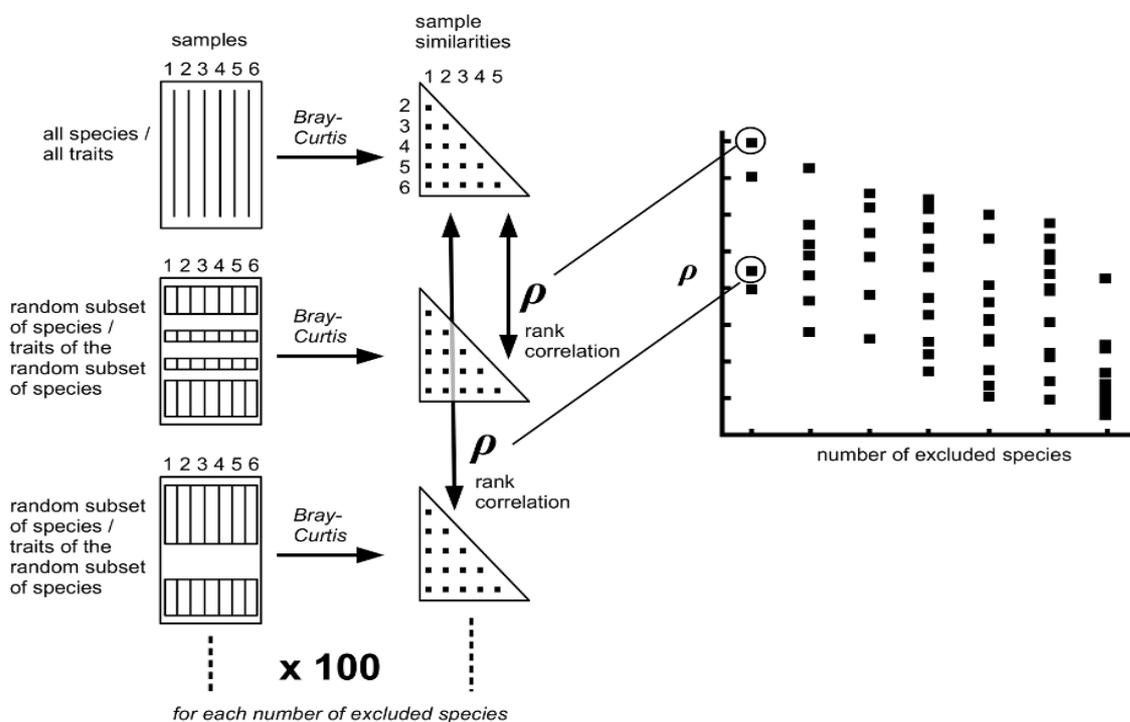


Figure 2: Schematic representation of the randomisation approach. The full species-by-sample matrix is converted into a triangular resemblance matrix by use of the Bray-Curtis coefficient. From the initial matrix, a random subset of species is selected and the new matrix is converted to a triangular resemblance matrix which is then compared to the resemblance matrix of the full set through a Spearman's correlation. The resulting correlation coefficient ρ is plotted against the number of excluded species. The process is repeated 100 times for each number of excluded species, until only one species is left. In parallel, this process is repeated for the traits-by-sample matrices, and the resulting correlation coefficients are plotted separately.

In parallel, for each subset of remaining species as determined above, a new traits-by-sample matrix was calculated, containing only traits expressed by the species in that subset. Thus, when a species was randomly excluded from the lagoon's inventory, certain traits were expressed with lower abundances in the new matrix (if remaining species express the same trait) or lost completely (if the lost species was unique in expressing a certain trait). For each of these new trait-by-sample matrices, comparisons with the initial matrix were performed as described above for the species-by-sample matrix. Since each species contributes to a different extent to the multivariate patterns, the removal of different species was expected to influence the pattern at varying degrees. Thus, species that do not play a major role in forming the initial pattern are expected not to change the pattern significantly when removed from the species pool, while others playing a more important role would cause a radical deterioration of the pattern when removed.

Again, it should be noted that these simulations neither reflect any particular real-world sequence of species loss nor do they take compensation effects of the remaining community or other ecosystem components into account. Instead, the results reflect the potential changes of patterns, capturing the best-case and worst-case scenario (note that *best* and *worst* here do not indicate any evaluation of the ecological status of the lagoon) into which real changes are likely to fall (see also section below “Limitations of this study”).

Comparison of simulation results

The results of the simulations for structural and functional patterns were compared by calculating four metrics for each of the two simulations: median (preferred here over the mean due to the very skewed distribution of the simulated values in many lagoons), minimum, maximum and range. These metrics summarise the distribution of the simulated similarity coefficients at each level of species richness. The distribution of these metrics, both for the structural and the functional patterns within each lagoon, were compared through a paired Wilcoxon test, assessing whether the two patterns are independent of each other. The degree of similarity between function and structure for each metric was compared with a Spearman's rank test.

All simulations and analyses were performed in R, version 2.15.1 (RCore team 2012). Data were stored in a MySQL database (<http://www.mysql.com>) and imported into R using the RMySQL library (James and DebRoy 2012). All data and R scripts can be downloaded from

figshare (doi:10.6084/m9.figshare.909347).

Results

The simulations of species loss showed, for each species richness level, the degree of similarity between the pattern of the initial dataset and the patterns obtained when a subset of species was removed. In all lagoons, an increasing loss of species caused the initial multivariate pattern to deteriorate (Figure 3). This could be observed for both the structural and the functional patterns. In most of the lagoons, the distributions of the simulated values of the structural and functional patterns showed a similar shape when visually compared to each other, but this shape was unique for each lagoon (Figure 3). The simulated values of both the structural and functional composition in the lagoons of Agiasma, Grado Marano and Margherita di Savoia showed a large variation and a heterogeneous distribution. In the lagoons of Logarou and Varna, the structural patterns showed bifurcations, whereas their functional patterns were more uniformly distributed. A uniform distribution of the simulated similarity values could also be observed for both the structural and the functional patterns in the lagoon of Grado Valle di Pesca.

Visual inspection of the simulated patterns also revealed that the variability of the similarity values around the median was large in Agiasma, Grado Marano and Margherita di Savoia, where the loss of a small percentage of species resulted in extremely variable changes of patterns, depending on the subset of excluded species. This variation was lower in Grado Valle di Pesca, Logarou and Varna, and was caused by different ratios of “redundant” and “dominant” species in each lagoon. The removal of the “redundant” species did not noticeably affect the overall pattern (maximum curve, $\rho > 0.95$). The percentage of redundant species was highest in Grado Marano, Agiasma and Varna, where, depending on their identity, more than 60% of the species could be removed before similarity values fell below $\rho > 0.95$. Likewise, in each lagoon, one or few dominant species were found whose removal immediately caused a sharp decrease in similarity values. However, despite the high similarity of the structural and functional patterns resulting from species loss, the species contributing most or least to the structural patterns were not always identical to those contributing most or least to the functional patterns (Tables 2, 3).

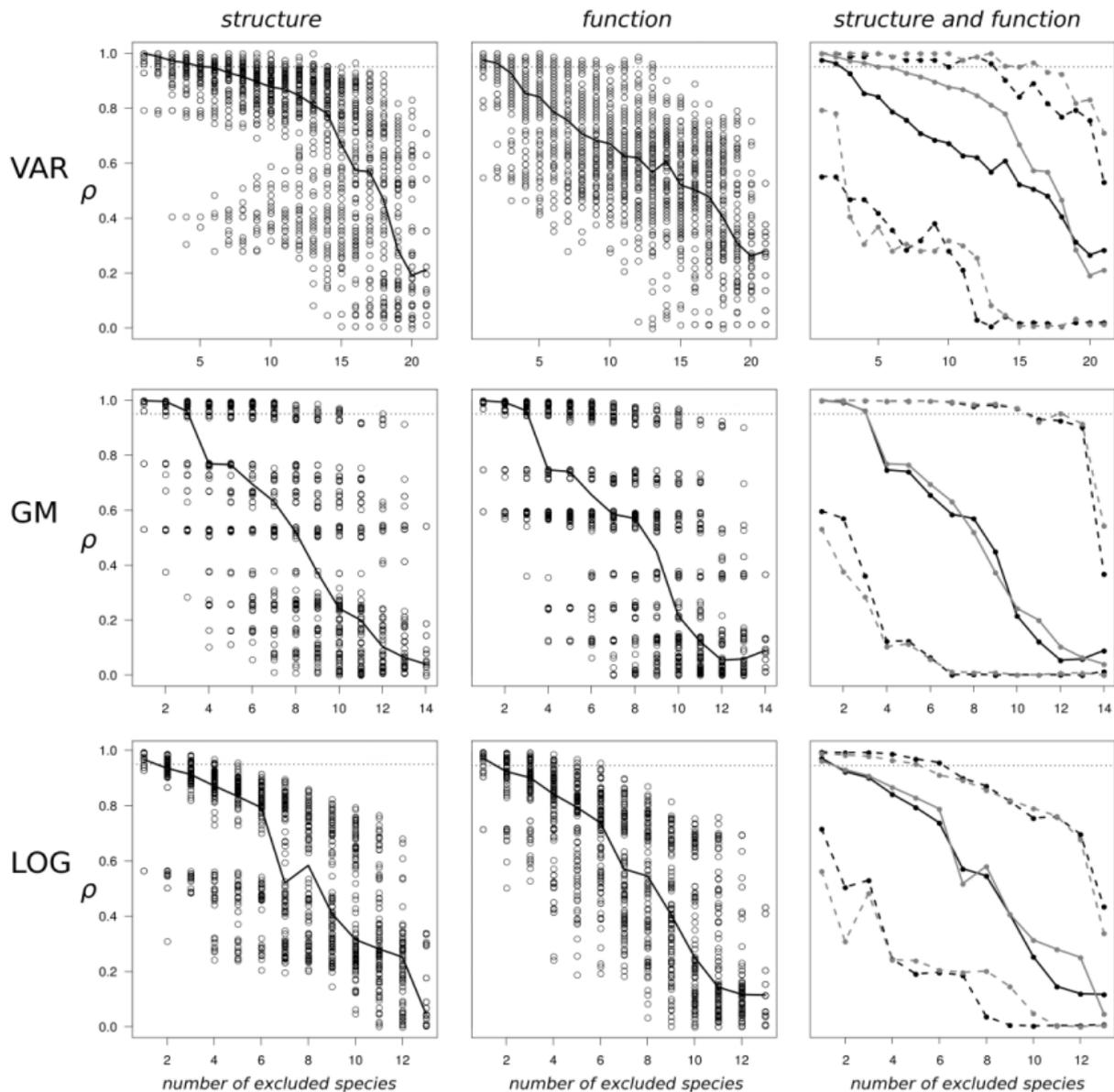


Figure 3 (continued on next page): Results of the simulations of random species loss for all lagoons. Abbreviations as in Figure 1. Left column: results for the species-by-sample matrices (structure), middle column: results for the traits-by-sample matrices (function), right column: minimum, maximum (dashed lines) and median (solid line) of both structure (black lines) and function (grey lines).

The results of the Spearman's rank correlation between the minimum, maximum, median and range showed, with a few exceptions, a high correlation between structural and functional patterns with increasing species loss, indicating that the change in community pattern had a significant effect on the functional patterns (Table 4).

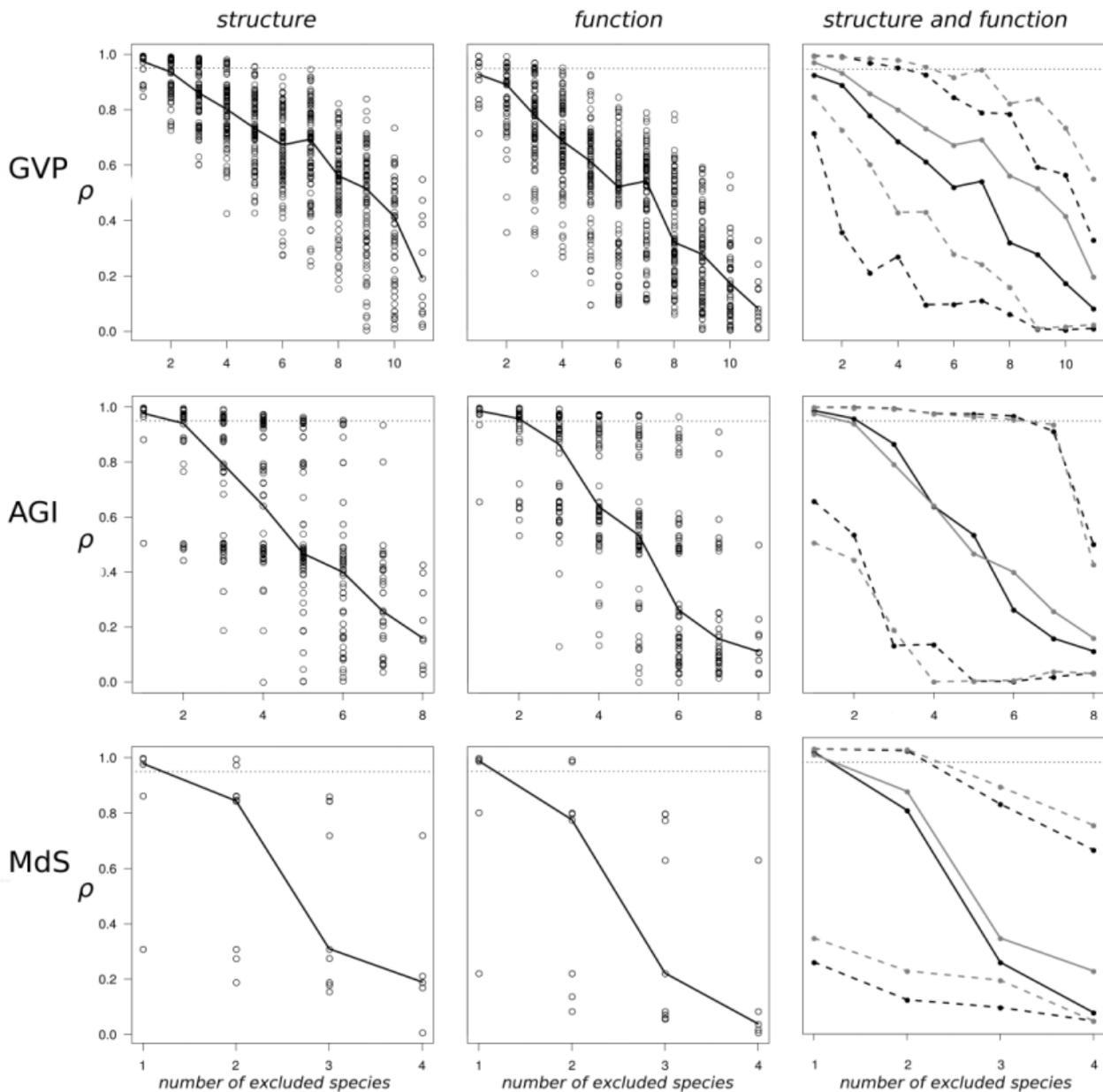


Figure 3 (continued): see previous page for legend

The median deterioration of both patterns was highly similar in each lagoon ($\rho > 0.98$), as was the number of redundant species (maximum curve, $\rho > 0.98$ in all lagoons except for Varna, $\rho = 0.81$). The minimum curve and the range of values showed a much lower correlation in most lagoons, an indication that species loss affected the magnitude of change in structure and function differently.

Table 2: Single species which, when removed, cause the strongest change in similarity to the full matrix within each lagoon. Spearman's correlation coefficient and the identity of the species are given both for the structural matrix and the functional matrix. Abbreviations as in Figure 1.

Lagoon	$\rho_{\text{structure}}$	species	ρ_{function}	species
AGI	0.51	<i>Nereis splendida</i>	0.66	<i>Nereis splendida</i>
GM	0.53	<i>Heteromastus filiformis</i>	0.6	<i>Heteromastus filiformis</i>
GVP	0.85	<i>Malacoceros fuliginosus</i>	0.71	<i>Streblospio shrubsolii</i>
LOG	0.57	<i>Nephtys hombergii</i>	0.72	<i>Nephtys hombergii</i>
MdS	0.31	<i>Hediste diversicolor</i>	0.22	<i>Hediste diversicolor</i>
VAR	0.79	<i>Streblospio shrubsolii</i>	0.55	<i>Streblospio shrubsolii</i>

Table 3: Single species which, when remaining as the only left species in the assemblage, creates the highest similarity to the full matrix within each lagoon. Spearman's correlation coefficient and the identity of the species are given both for the structural matrix and the functional matrix. Abbreviations as in Figure 1.

Lagoon	$\rho_{\text{structure}}$	species	ρ_{function}	species
AGI	0.42	<i>Nereis splendida</i>	0.5	<i>Capitella capitata capitata</i>
GM	0.54	<i>Heteromastus filiformis</i>	0.36	<i>Heteromastus filiformis</i>
GVP	0.55	<i>Malacoceros fuliginosus</i>	0.32	<i>Malacoceros fuliginosus</i>
LOG	0.34	<i>Hydroides dianthus</i>	0.43	<i>Nephtys hombergii</i>
MdS	0.72	<i>Hediste diversicolor</i>	0.63	<i>Hediste diversicolor</i>
VAR	0.72	<i>Melinna palmata</i>	0.53	<i>Alitta succinea</i>

The paired Wilcoxon test showed that significant differences ($p < 0.05$) between function and structure could mainly be detected in the lagoons of Grado Valle di Pesca (for the maximum, minimum and median values) and Varna (maximum and median values) (Table 5). The maximum values in Logarou and the range in Grado Marano also showed significant differences, but in no other case could differences in any of the metrics be detected. In Grado Valle di Pesca, this difference was manifested through a “shift” of the functional values towards lower similarities – the functional patterns deteriorated at a faster rate than the structural patterns. A similar trend could be observed in Varna for the maximum values after a certain

percentage of species was lost. The smallest species group identified as sufficient to reproduce a structural pattern of $\rho > 0.95$ was only able to reproduce a functional pattern of $\rho = 0.93$ in Grado Valle di Pesca and $\rho = 0.89$ in Varna. Thus, in these two lagoons, the “surviving” species could still reproduce a structural pattern highly similar to the pattern of the total community, whereas the functional pattern could not be maintained any more at a similar level.

Table 4: Results of the Spearman correlation between structure and function for each lagoon, for the maximum, minimum, median and range of the simulation results. Abbreviations as in Figure 1.

Lagoon	max		min		median		range	
	ρ	p value						
AGI	0.994	<0.001	0.6	0.132	1	<0.001	0.690	0.07
GM	0.991	<0.001	0.77	0.001	0.982	<0.001	0.978	<0.001
GVP	0.982	<0.001	0.909	<0.001	1	<0.001	-0.036	0.924
OG	0.989	<0.001	0.94	<0.001	0.995	<0.001	0.511	<0.001
MdS ^a	1	0.083	1	0.083	1	0.08333	0.2	0.917
VAR	0.81	<0.001	0.893	<0.001	0.999	<0.001	0.794	<0.001

^a Non-significant results are caused by the low species number ($s=5$)

Table 5: Results of the paired Wilcoxon test between structure and function for each lagoon, for the maximum, minimum, median and range of the simulation results. Abbreviations as in Figure 1.

Lagoon	max		min		median		range	
	V	p value	V	p value	V	p value	V	p value
AGI	7	0.148	13	0.547	20	0.844	22	0.641
GM	77	0.135	32	0.217	67	0.391	88	0.025
GVP	63	0.005	66	<0.001	66	<0.001	22	0.365
LOG	14	0.027	46	0.839	67	0.147	31	0.334
MdS	10	0.125	10	0.125	9	0.125	3	0.625
VAR	129	<0.001	120	0.876	213	<0.001	156	0.168

Discussion

Species richness, dominance and redundancy

The simulations of the present study showed that in all lagoons an increasing loss of species resulted in an increasing deterioration of both the structural and functional multivariate patterns, and functions were lost at a similar rate to species loss. Theory predicts that species-rich systems are more stable (Lawton and Brown 1993, Loreau et al. 2001, McCann 2000, Petchey 2000, Tilman et al. 1997b), since there is a greater probability that important functions

are performed by multiple species, and functions of lost species can potentially be performed by the remaining species. This provides the system with a certain resilience towards species loss and thus loss of function, a concept known as the “insurance hypothesis” (Yachi and Loreau 1999). In species-poor systems on the other hand, the trait pool is already reduced and since many functions are only performed by a single species, even the loss of one species can have extreme effects if it leads to the complete extinction of certain functions from the system.

The results of this study do, however, not fully confirm the importance of species richness for ecosystem stability. In fact, even in the most impoverished lagoon, Margharita di Savoia (5 species), redundancy of species could be observed (Figure 3, maximum curve). The second most species-rich lagoon, Grado Marano (15 species), showed a large range of possible scenarios following species loss: a large percentage of species were redundant (maximum curve) – their loss affected the existing pattern only minimally, whereas other species played an overly important role in the system and their loss had dramatic effects on the functional composition of the system (Figure 3, minimum curve). In contrast, Grado Valle di Pesca and Logarou, with a similar number of species (12 and 14 species, respectively) had a smaller potential for extreme changes (less variation around the median). It appears that in these already impoverished systems, species richness alone cannot be used to explain the effects of species loss on the functioning of the system, and other factors, such as species identity and community composition are likely to play a role as well. In fact, a number of studies have identified species composition to be more important than species richness (Symstad et al. 1998, Tilman et al. 1997a, Tilman et al. 1997b). Smith and Knapp (2003) have found that productivity in grassland experiments increased with decreasing species richness, with dominant species compensating for the loss of rare species and even being able to increase the system's short-term resilience. Dominant species are known to have a strong influence on functional diversity (Loreau et al. 2001, Mistri et al. 2001, Smith and Knapp 2003, Symstad et al. 1998), an effect which is caused by strong interspecific competition (Petchey 2000). This phenomenon was recognised by Walker (1992) and led to the concept of “driver” and “passenger” species. In the present study, all lagoons showed, to different extents, the existence of these “driver” and “passenger” species (the effect of their loss are represented by the minimum and maximum curves in Figure 3). The redundant species had a negligible effect on the change of the functional patterns when removed, whereas the loss of dominant species significantly changed the patterns. On the other hand, dominant species provided stability to the system if they were

among the “survivors” – changes were less drastic. In Margherita di Savoia and Varna the distribution of a single remaining species could still produce a pattern of $\rho > 0.5$ similarity, in Grado Marano and Agiasma, two species were able to almost fully reproduce the pattern of the total community ($\rho > 0.9$). In Grado Valle di Pesca and Logarou, on the other hand, no single species was able to preserve the functional patterns, an indication that complementarity rather than dominance is the main structuring factor for these communities (Loreau et al. 2001).

These differences in variance across the different systems can furthermore be explained by the initial composition of the system. The lagoons of Agiasma, Grado Marano and Margherita di Savoia which are characterised by environmental gradients (eutrophication, chemical pollution and salinity, respectively) showed an uneven distribution of species across the samples, with some samples containing only a single species. These disturbed systems (either by natural or anthropogenic factors) showed a larger variation concerning the effects of species loss on the structural and functional patterns: the loss of certain species either affected the system not at all, or it induced extreme changes.

More balanced systems such as Grado Valle di Pesca, Logarou and Varna, where species were distributed more evenly, showed a less extreme variation around the median and can therefore be considered to be more robust and less likely to suffer extreme changes. The loss of a certain species from two different communities with similar species composition and richness (e.g. Grado Marano and Grado Valle di Pesca) might thus provoke unpredictable ecosystem changes which depend on factors such as community composition, trait expression and environmental conditions (Symstad et al. 1998, Petchey and Gaston 2002, Wilmers et al. 2002). These results support the hypothesis that the stability of a system indeed depends more on the identity and frequency of occurrence of the species that are removed or remain in the community than on species richness alone. Richness may play a greater role in less impacted communities characterised by complementarity, whereas even in naturally disturbed communities, the dominance of one or few species is the major driver for the observed patterns. This is exemplified by the lagoon of Varna, which despite being subjected to environmental stressors such as pollution, showed less potential for extreme changes than similarly stressed lagoons with lower species richness, since the remaining species can compensate for the losses, at least up to a certain point. The findings that species composition and identity seem to determine the effects of species loss on the functional patterns, support the idiosyncratic hypothesis (Naeem et al. 1995), which states that ecosystem processes change

when diversity changes, but that the extent of the changes is unpredictable and context-dependent (Cardinale et al. 2006, Hooper et al. 2005, Naeem and Wright 2003, Naeem et al. 1995). Albeit postulated for non-random species loss (Naeem et al. 2002), the hypothesis may also explain the variability around an average rate of functional loss in random simulations. This variability can provide important information about the potential stability of a system (Fonseca and Ganade 2001). It appears here that the ability to predict changes in the system inflicted by species loss is greater in more balanced systems, whereas already disturbed systems show a much larger degree of unpredictability. This is of direct practical importance for conservation decisions, since systems with a large variation in effects of extinction are much more difficult to manage (Petchey 2000).

Differences between function and structure

The simulated species loss had a very similar effect on both the structural and the functional patterns in all lagoons except Grado Valle di Pesca and Varna, where the loss of species caused a stronger deterioration of the functional pattern than that of the structural pattern. This might be attributed to the complementarity observed in these lagoons: individual species' contribution to functioning had a greater weight in these systems, and thus the loss of unique traits has a greater impact. However, despite the high similarity of the structural and functional patterns in most of the lagoons, the identity of the species contributing most to either pattern was not always the same (Tables 2, 3). It appears that in communities where dominance of a few species is the main driver for ecosystem functioning, structural stability is to be indeed a proxy for functional stability, and even the identity of the species dominating either patterns is identical. If these findings can be confirmed for other ecosystems, they might be of great use for the development of rapid assessment tools required for ecosystem management (Allan et al. 2006). In systems where complementarity shared among many species was the main driver for functioning, the patterns between functioning and structure diverged, and different species emerged as the “driver” species for either pattern. Although there is lots of evidence for structural redundancy in marine macrobenthic assemblages (Clarke and Warwick 1998, Olsgard et al. 1997, 1998), it has never formally been tested whether these structurally redundant species are also functionally interchangeable. If the results of this study are shown to be constant, relying on structural redundancy to assess ecosystems resilience might in fact overestimate the latter and misjudge the importance of specific species for ecosystem functioning.

Limitations of the study

As with any simulation, this study required a simplification and abstraction of natural processes, resulting in being a rather hypothetical approximation of potential real scenarios. One of the major simplifications is, of course, that the model does not take compensation of species loss by other community members or other ecosystem components (e.g. other macrobenthic species, meiobenthos, microbes) into account. In real-world biological communities, the loss of a certain species can free other species from competition, resulting in their proliferation or even dominance, which in turn might have an impact on other species and thus the whole community composition. The effects of such a cascading scenario (Petchey et al. 2004) are unpredictable, because changes in the abundance or biomass patterns can strongly affect ecosystem functioning (Larsen et al. 2005), and remaining species, depending on their limitations concerning reproduction and dispersal, may or may not be able to compensate for the lost functions (Larsen et al. 2005, Snelgrove 1998, Solan et al. 2004). However, in systems with low species richness and thus an already reduced species pool, certain functions are completely lost when a species is lost, and even numerical compensation of the surviving species will not restore this function. For numerical compensation of reduced functions, the ecosystem needs therefore to have a certain level of functional redundancy, and the “redundant” species need furthermore to have the reproductive capability to take over the role of the lost species (Snelgrove 1998).

Furthermore, our simulations do not take any recolonisation of the system by other species into account. Coastal lagoons are in almost constant exchange with the adjacent sea, from which their species pool is recruited. Depending on the environmental conditions within the lagoon, an impoverished community can be recolonised in a relatively short time by species immigrating from the sea – a process which in some lagoons takes place on a regular seasonal basis (Barnes 1980). Such recolonisation might even introduce additional traits to the existing pool (Mayfield et al. 2010), thus adding different ecosystem functions and altering the processes within the environment. Modelling these dynamics is, however, not within the scope of this study. On the other hand, coastal lagoons set strong environmental barriers (e.g. salinity) which have a robust selective power to species distributions (Bamber et al. 1992, Barnes 1988), and only species with appropriate attributes are able to survive and perhaps become dominant under such severe environmental conditions. The loss of these species would, therefore, be a result of aggravated environmental conditions. Consequently, the assumption made in this

study that lost species are unlikely to be replaced by re-entry from the sea is considered to be fairly realistic.

Finally, the dominant species in the present datasets are dominant for a certain reason – environmental conditions favour their proliferation or prevent their extinction. In reality, these species are likely to have a much smaller chance of going extinct, whereas the loss of a rare species is much more likely. Other extinction scenarios exist likewise – large species being lost first, or species in a higher level in the food web being lost first (Pimm et al. 1988). An ordered extinction model simulating species loss directed by factors such as abundance, frequency, body size or trophic level could address this question (Duffy 2003, Petchey and Gaston 2002, Solan et al. 2004), but for such an approach a certain level of species richness is required. The naturally impoverished lagoonal communities investigated in the present study does not permit this approach. On the other hand, the low species richness made it possible to test a high percentage of possible species combinations and their effect on the functional patterns (Vitousek and Hooper, 1993, O'Connor and Crowe 2005). The results therefore provide a good approximation for the extent of possible changes occurring in these lagoons in cases of species loss.

Conclusions

This study investigates the potential effects of species loss on the community structure and functioning in coastal lagoons. In all studied systems, the loss of species has a deteriorating effect on both the patterns of the community structure and the functional patterns expressed by this community. In disturbed systems dominated by few species, both structural and functional patterns change in almost identical ways, rendering structural redundancy a surrogate of functional redundancy. In less disturbed systems, however, species loss affects community structure differently from its functionality, indicating that relying on structural redundancy as a measure of ecosystem resilience might lead to an overestimation of the latter and thus to false conclusions.

Furthermore, the results show that the effects of species loss on functional patterns cannot be explained through simple stochastic factors such as species richness alone. Instead, the initial community composition and the identity of the lost species seems to play a major role in determining the response of the ecosystem. The results of this study indicate that species richness seems to play a greater role in the less disturbed communities characterised by complementarity, whereas in the severely disturbed ones, the dominance of one or few species is

the major driver for the observed patterns, lending support to the paradigm of “drivers” and “passengers” as described by Walker (1992). These systems react more extremely to the loss of species concerning their functional patterns, making the prediction of changes very difficult at best. Less impacted lagoons, on the other hand, show indications of the rivet model: here, most species provide unique functions to the system and their loss is more likely to have an effect, but on average this effect is less dramatic and the system more stable.

These findings therefore highlight the importance of considering the idiosyncratic effects of species loss on ecosystem function as well as the necessity to take community composition and species extinction risks into account when taking decisions at a managerial level.

References

- Allan IJ, Vrana B, Greenwood, R, Mills GA, Roig B, Gonzales C (2006) A “toolbox” for biological and chemical monitoring requirements for the European Union's Water Framework Directive. *Talanta* 69: 302–322. doi:10.1016/j.talanta.2005.09.043
- Arvanitidis C, Hatzigeorgiou, G, Koutsoubas, D, Dounas, C, Eleftheriou, A, Koulouri, P (2005) Mediterranean lagoons revisited: weakness and efficiency of the rapid biodiversity assessment technique in severely fluctuating environment. *Biodiversity Conservation* 14: 2347–2359. doi:10.1007/s10531-004-1668-x
- Bamber RN, Batten SD, Shearer M, Bridgwater ND (1992) On the ecology of brackish water lagoons in Great Britain. *Estuarine, Coastal and Shelf Science* 26: 309–318. doi:10.1016/0272-7714(88)90067-4
- Barnes RSK (1980) *Coastal Lagoons*, Cambridge University Press, Cambridge, 120pp.
- Barnes RSK (1988) The faunas of land-locked lagoons: Chance differences and the problems of dispersal. *Estuarine, Coastal and Shelf Science* 26: 308–318.
- Barnes RSK (1991) European estuaries and lagoons: a personal overview of problems and possibilities for conservation and management. *Aquatic Conservation: Marine and Freshwater Ecosystems* 1, 79–87. doi:10.1002/aqc.3270010107
- Barbone E, Rosati I, Pinna M, Basset A (2007) Taxonomic and dimensional structure of benthic macroinvertebrate guilds in the Margherita di Savoia Salt Pans (Italy). *Transitional Waters Bulletin* 4: 21–31.
- Basset A, Sabetta L, Sangiorgio F, Pinna M, Migoni D, Fanizzi F, Barbone E, Galuppo N,

- Umani SF, Reizopoulou S, Nicolaidou A, Arvanitidis C, Moncheva S, Trajanova A, Georgescu L, Beqiraj S (2008) Biodiversity conservation in Mediterranean and Black Sea lagoons: a trait-oriented approach to benthic invertebrate guilds. *Aquatic Conservation: Marine and Freshwater Ecosystems* 15: 4–15. doi:10.1002/aqc.979
- Basset A, Barbone E, Borja A, Bruce S, Pinna M, Quintana XD, Reizopoulou S, Rosati I, Simboura N (2012) A benthic macroinvertebrate size spectra index for implementing the Water Framework Directive in coastal lagoons in Mediterranean and Black Sea. *Ecological Indicators* 12: 72–83. doi:10.1016/j.ecolind.2011.06.012
- Basset A, Barbone E, Elliot M, Li BL, Jorgensen SE, Lucena-Moya P, Pardo I, Mouillot D (2013) A unifying approach to understanding transitional waters: Fundamental properties emerging from ecotone ecosystems. *Estuarine, Coastal and Shelf Science* 132: 5–16. doi:10.1016/j.ecss.2012.04.012
- Boero F, Bonsdorff E (2007) A conceptual framework for marine biodiversity and ecosystem functioning. *Marine Ecology* 28: 134–145. doi:10.1111/j.1439-0485.2007.00171.x
- Bray JR, Curtis JT (1957) An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs* 27: 325–349. doi:10.2307/1942268
- Bremner J (2008) Species' traits and ecological functioning in marine conservation and management. *Journal of Experimental Marine Biology and Ecology* 366: 37–47. doi:10.1016/j.jembe.2008.07.007
- Bremner J, Rogers S, Frid CLJ (2003) Assessing functional diversity in marine benthic ecosystems: a comparison of approaches. *Marine Ecology Progress Series* 254: 11–25. doi:10.3354/meps254011
- Bremner, J, Rogers SI, Frid CLJ (2006) Matching biological traits to environmental conditions in marine benthic ecosystems. *Journal of Marine Systems* 60: 302–316. doi:10.1016/j.jmarsys.2006.02.004
- Cardinale BJ, Srivastava DS, Duffy JE, Wright JP, Downing AL, Sankaran M, Jouseau C (2006) Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* 443, 989–992. doi:10.1038/nature05202
- Clarke KR, Warwick RM (1998) Quantifying structural redundancy in ecological communities. *Oecologia* 113: 278–289. doi:10.1007/s004420050379
- Covich AP, Austen MCV, Bärlocher F, Chauvet E, Cardinale BJ, Biles CL, Inchausti P, Dangles O, Solan M, Gessner MO, Stutzner B, Moss B (2004) The Role of Biodiversity in the

- Functioning of Freshwater and Marine Benthic Ecosystems. *Bioscience* 54: 767–775.
doi:10.1641/0006-3568(2004)054[0767:TROBIT]2.0.CO;2
- Danovaro R, Pusceddu A (2007) Ecomanagement of biodiversity and ecosystem functioning in the Mediterranean Sea: concerns and strategies. *Chemistry and Ecology* 23: 347–360.
doi:10.1080/02757540701653384
- Díaz S, Symstad AJ, Chapin FS III, Wardle DA, Huenneke LF (2003) Functional diversity revealed by removal experiments. *Trends in Ecology and Evolution* 18: 140–146.
doi:10.1016/S0169-5347(03)00007-7
- Duffy JE (2003) Biodiversity loss, trophic skew and ecosystem functioning. *Ecology Letters* 6: 680–687. doi:10.1046/j.1461-0248.2003.00494.x
- Ehrlich P, Ehrlich A (1981) *The causes and consequences of the disappearance of species*, Random House, New York, 384 pp.
- Fauchald, K (1977) *The polychaete worms. Definitions and keys to the orders, families and genera*, Natural History Museum of Los Angeles county, Los Angeles, 198pp.
- Faulwetter S, Markantonatou V, Pavludi C, Papageorgiou N, Keklikoglou K, Chatzinikolaou E, Pafilis E, Dailianis T, Chatzigeorgiou G, Vasileiadou K, Fanini L, Koulouri P, Arvanitidis C (2014) *Polytraits*: A database on biological traits of marine polychaetes. *Biodiversity Data Journal* 2:e1024. doi: 10.3897/BDJ.2.e1024
- Fonseca CR, Ganade G (2001) Species functional redundancy, random extinctions and the stability of ecosystems. *Journal of Ecology* 89: 118–125.
doi:10.1046/j.1365-2745.2001.00528.x
- Gamfeldt L, Hillebrand H (2008) Biodiversity Effects on Aquatic Ecosystem Functioning - Maturation of a New Paradigm. *International Review of Hydrobiology*. 93, 550–564. doi: 10.1002/iroh.200711022
- Giangrande A, Licciano M, Musco L (2005) Polychaetes as environmental indicators revisited. *Marine Pollution Bulletin* 50: 1153–1162. doi:10.1016/j.marpolbul.2005.08.003
- Hewitt JE, Thrush SF, Dayton PD (2008) Habitat variation, species diversity and ecological functioning in a marine system. *Journal of Experimental Marine Biology and Ecology* 366: 116–122. doi:10.1016/j.jembe.2008.07.016
- Hooper DU, Chapin FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, Schmid B, Setälä H, Symstad AJ, Vandermeer J, Wardle DA (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge.

- Ecological Monographs 75: 3–35. doi:10.1890/04-0922
- Hooper DU, Vitousek PM (1997) The Effects of Plant Composition and Diversity on Ecosystem Processes. *Science* 277: 1302–1305. doi:10.1126/science.277.5330.1302
- Hutchings PA (1998) Biodiversity and functioning of polychaetes in benthic sediments. *Biodiversity and Conservation* 7: 1133–1145. doi:10.1023/A:1008871430178
- Ives AR, Carpenter SR (2007) Stability and diversity of ecosystems. *Science* 317: 58–62. doi:10.1126/science.1133258
- James DA, DebRoy S (2012) RMySQL: R interface to the MySQL database. R package version 0.9-3. <http://CRAN.R-project.org/package=RMySQL>. Last accessed: 14 April 2013.
- Larsen TH, Williams NM, Kremen C (2005) Extinction order and altered community structure rapidly disrupt ecosystem functioning. *Ecology Letters* 8, 538–547. doi:10.1111/j.1461-0248.2005.00749.x
- Lawton JH, Brown VK (1993) Redundancy in Ecosystems. In: Schulze ED, Mooney HA (Eds) *Biodiversity and Ecosystem Function*, Springer, Berlin Heidelberg, 255–270.
- Loreau M, Naeem S, Inchausti P, Bengtsson J, Grime JP, Hector A, Hooper DU, Huston MA, Raffaelli DG, Schmid B, Tilman D, Wardle DA (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* 294: 804–808. doi:10.1126/science.1064088
- Marchini, A, Munari C, Mistri M (2008) Functions and ecological status of eight Italian lagoons examined using biological traits analysis (BTA). *Marine Pollution Bulletin* 56: 1076–1085. doi: 10.1016/j.marpolbul.2008.03.027
- Mayfield MM, Bonser SP, Morgan JW, Aubin I, McNamara S, Vesik PA (2010) What does species richness tell us about functional trait diversity? Predictions and evidence for responses of species and functional trait diversity to land-use change. *Global Ecology and Biogeography* 19: 423–431. doi:10.1111/j.1466-8238.2010.00532.x
- McCann KS (2000) The diversity-stability debate. *Nature* 405, 228–233. doi:10.1038/35012234
- Mistri M, Fano E, Rossi R (2001) Redundancy of macrobenthos from lagoonal habitats in the Adriatic Sea. *Marine Ecology Progress Series* 215: 289–296. doi:10.3354/meps215289
- Mouillot D, Laune J, Tomasini JA, Aliaume C, Brehmer P, Dutrieux E, Do Chi T (2005) Assessment of Coastal Lagoon Quality with Taxonomic Diversity Indices of Fish, Zoobenthos and Macrophyte Communities. *Hydrobiologia* 550, 121–130. doi:10.1007/s10750-005-4368-y

- Naeem S (2006) Expanding scales in biodiversity-based research: challenges and solutions for marine systems. *Marine Ecology Progress Series* 311: 273–283. doi:10.3354/meps311273
- Naeem S, Bunker DE, Hector A, Loreau M, Perrings C (Eds) 2009. *Biodiversity, Ecosystem Functioning, and Human Wellbeing. An Ecological and Economic Perspective*, Oxford University Press, Oxford, 388pp.
- Naeem, S, Loreau, M, Inchausti, P. 2002. Biodiversity and ecosystem functioning : the emergence of a synthetic ecological framework. In: Naeem, S, Loreau, M, Inchausti P. (Eds) *Biodiversity and ecosystem functioning: synthesis and perspectives*, Oxford University Press, Oxford, 3–17.
- Naeem S, Thompson LJ, Lawler SP, Lawton JH, Woodfin RM (1995) Empirical Evidence that Declining Species Diversity May Alter the Performance of Terrestrial Ecosystems. *Philosophical Transactions of the Royal Society B* 347: 249–262. doi:10.1098/rstb.1995.0025
- Naeem S, Wright JP (2003) Disentangling biodiversity effects on ecosystem functioning: deriving solutions to a seemingly insurmountable problem. *Ecology Letters* 6: 567–579. doi:10.1046/j.1461-0248.2003.00471.x
- O'Connor, NE, Crowe TP (2005) Biodiversity Loss and Ecosystem Functioning: Distinguishing Between Number and Identity of Species. *Ecology* 86: 1783–1796. doi:10.1890/04-1172
- Olgard F, Somerfield PJ, Carr MR (1997) Relationships between taxonomic resolution and data transformations in analyses of a macrobenthic community along an established pollution gradient. *Marine Ecology Progress Series* 149: 173–181. doi:10.3354/meps149173
- Olgard F, Somerfield PJ, Carr MR (1998) Relationships between taxonomic resolution, macrobenthic community patterns and disturbance. *Marine Ecology Progress Series* 172: 25–36. doi:10.3354/meps172025
- Petchey OL (2000) Species Diversity, Species Extinction, and Ecosystem Function. *The American Naturalist* 155: 696–702. doi:10.1086/303352
- Petchey OL, Downing AL, Mittelbach GG, Persson L, Steiner, CF, Warren PH, Woodward G (2004) Species loss and the structure and functioning of multitrophic aquatic systems. *Oikos* 104: 467–478. doi:10.1111/j.0030-1299.2004.13257.x
- Petchey OL, Gaston KJ (2002) Extinction and the loss of functional diversity. *Proceedings of the Royal Society of London Series B* 269: 1721–1727. doi:10.1098/rspb.2002.2073
- Pimm SL, Jones HL, Diamond JM (1988) On the risk of extinction. *The American Naturalist*

- 6: 757–785.
- Reizopoulou S, Nicolaidou A (2004) Benthic diversity of coastal brackish-water lagoons in western Greece. *Aquatic Conservation: Marine and Freshwater Ecosystems* 14: S93–S102. doi:10.1002/aqc.653
- Reizopoulou S, Simboura N, Barbone E, Aleffi F, Basset A, Nicolaidou A (in press) Biodiversity in transitional waters: steeper ecotone, lower diversity. *Marine Ecology*. doi:10.1111/maec.12121
- R Core Team 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Sangiorgio F, Basset A, Pinna M, Sabetta L, Abbiati M, Ponti M, Minocci M, Orfanidis S, Nicolaidou A, Moncheva S, Trayanova A, Georgescu L, Dragan S, Beqiraj S, Koutsoubas D, Evagelopoulos A, Reizopoulou S (2008) Environmental factors affecting *Phragmites australis* litter decomposition in Mediterranean and Black Sea transitional waters. *Aquatic Conservation: Marine and Freshwater Ecosystems* 26: 16–26. doi:10.1002/aqc.955
- Sigala K, Reizopoulou S, Basset A, Nicolaidou A (2012) Functional diversity in three Mediterranean transitional water ecosystems. *Estuarine, Coastal and Shelf Science* 110: 202–209. doi:10.1016/j.ecss.2012.06.002
- Smith MD, Knapp AK (2003) Dominant species maintain ecosystem function with non-random species loss. *Ecology Letters* 6: 509–517. doi:10.1046/j.1461-0248.2003.00454.x
- Snelgrove PVR (1998) The biodiversity of macrofaunal organisms in marine sediments. *Biodiversity and Conservation* 7: 1123–1132. doi:10.1023/A:1008867313340
- Solan M, Cardinale BJ, Downing AL, Engelhardt KAM, Ruesink JL, Srivastava DS (2004) Extinction and ecosystem function in the marine benthos. *Science* 306: 1177–1180. doi:10.1126/science.1103960
- Solan M, Raffaelli DG, Paterson DM, White, PCL, Pierce GJ (Eds) (2006) Marine biodiversity and ecosystem function: empirical approaches and future research needs. *Marine Ecology Progress Series* 311: 175–309.
- Srivastava DS, Vellend M (2005) Biodiversity-Ecosystem function research: Is It Relevant to Conservation? *Annual Review of Ecology, Evolution, and Systematics* 36, 267–294. doi:10.1146/annurev.ecolsys.36.102003.152636
- Symstad AJ, Chapin SFI, Wall DH, Gross KL, Huenneke LF, Mittelbach GG, Peters DP, Tilman

-
- D (2003) Long-Term and Large-Scale Perspectives on the Relationship between Biodiversity and Ecosystem Functioning. *Bioscience* 53: 89–98.
doi:10.1641/0006-3568(2003)053[0089:LTALSP]2.0.CO;2
- Symstad AJ, Tilman D, Willson J, Knops JMH (1998) Species loss and ecosystem functioning: effect of species identity and community composition. *Oikos* 81: 389–397.
- Tilman D, Knops J, Wedin D, Reich PB, Ritchie M, Siemann E (1997a) The influence of functional diversity and composition on ecosystem processes. *Science* 277: 1300–1302.
doi:10.1126/science.277.5330.1300
- Tilman D, Lehman C, Thomson KT (1997b) Plant diversity and ecosystem productivity: Theoretical considerations. *Proceedings of the National Academy of Sciences* 94: 1857–1861.
- Törnroos A, Bonsdorff E (2012) Developing the multitrait concept for functional diversity: lessons from a system rich in functions but poor in species. *Ecological Applications* 22, 2221–2236. doi:10.1890/11-2042.1
- Tyler EHM, Somerfield PJ, Vanden Berghe E, Bremner J, Jackson E, Langmead O, Palomares MLD, Webb TJ (2011) Extensive gaps and biases in our knowledge of a well-known fauna: implications for integrating biological traits into macroecology. *Global Ecology and Biogeography* 21: 922–934. doi:10.1111/j.1466-8238.2011.00726.x
- Vitousek PM, Hooper DU (1993) Biological diversity and terrestrial ecosystem biogeochemistry. In Schulze ED, Mooney HA (Eds) *Biodiversity and ecosystem function*. Springer, New York, 3–14.
- Walker BH (1992) Biodiversity and ecological redundancy. *Conservation Biology* 6: 18–23.
doi:10.1046/j.1523-1739.1992.610018.x
- Wilmers CC, Sinha S, Brede M (2002) Examining the effects of species richness on community stability: an assembly model approach. *Oikos* 99: 363–367.
doi:10.1034/j.1600-0706.2002.990218.x
- Yachi S, Loreau M (1999) Biodiversity and ecosystem functioning in a fluctuating environment: the insurance hypothesis. *Proceedings of the National Academy of Sciences* 96: 1463–1468.

Chapter 6

Discussion and Conclusions

Discussion

Data form the basis of scientific progress: they provide measurable evidence of the world around us and lead to the postulation and testing of new hypotheses. Fast computers, high-speed internet and the increasing availability of storage media have resulted in enormous volumes of data being readily available for processing and analyses by powerful computers. However, whereas in certain scientific disciplines (e.g. earth and climate changes, oceanography) raw data are readily made available to the public, the biodiversity community remains rather reluctant to share their data publicly (Smith 2009). This is a known phenomenon and a range of explanations as well as possible remedies to change this behaviour have been proposed (e.g. Chavan and Ingwersen 2009, Huang and Qiao 2011, Reichman et al. 2011, Tenopir et al. 2011). Given that the field has undergone radical technological changes during the last two decades, it is understandable that researchers' every-day practice needs some time to adapt to these changes. Nevertheless, the attitude towards data sharing is only slowly improving.

This thesis demonstrates how scientific data management and publication can be embedded into every-day research by using existing tools and methods, provides cases which show how these technological advancements can be used for the creation of new scientific knowledge, and identifies challenges and gaps in current practices in the discipline.

Incentives and obstacles to data sharing

Modern science relies more than ever on modern methods of producing, sharing and reusing data (Schnase, 2000; Tenopir et al., 2011). Daily scientific practice shows, however, that although most research is conducted digitally, most scientific end-products are simple PDF files, whereas the original data from which the results are derived often disappear after the publication of the paper (Parr et al. 2011). In biodiversity research, sound data management plans which include the submission of the research data to public repositories are still rather exceptions than a rule. This has to do with a variety of factors (Costello 2009, Enke et al. 2012), some of which stem from the current scientific culture and can be detected in other disciplines as well; others may originate from the specific nature of biodiversity research. Often, a lack of formal regulations for data sharing – either by funding agencies or by scientific publishers – is named as a major cause for the reluctance to share data. Specifically in the field of biodiversity, another argument often voiced is that scientists are reluctant to share their data due to the

especially high effort required to obtain them (hard work in the field, time-consuming manual identifications of specimens in the laboratory) and fear that others might receive undue credit if the data are reused. These arguments undoubtedly hold true to a certain extent, and are increasingly being addressed by emerging legislations¹, journals' requirements (Stodden et al. 2013), initiatives which promote good practice² as well as by the increasing availability of repositories³ that allow scientists to easily publicise their data and receive a citable Digital Object Identifier (DOI) in return.

Yet, the real problem likely lies elsewhere and is not remedied by these efforts: scientists have no or little direct benefit from sharing data. To properly publish their dataset, they need to invest time and effort which could otherwise be invested in activities that result in positive indicators for their scientific career: the publication of papers and the acquisition of projects and thus monetary funds (Smith 2009, Smith et al. 2011, Enke et al. 2012). In addition, data management and data publishing is not an easy task if it is to be performed correctly. The simple submission of the raw data file(s) to an online repository such as figshare or DataDryad is a start concerning the public availability of data, but ideally datasets should be quality-controlled, formatted according to community standards (e.g. Darwin Core), annotated with appropriate metadata and published through recognised channels. Given that a plethora of databases, standards and guidelines for data sharing and formatting exist, researchers are often unable to keep up these technical requirements. New technologies require learning to use them properly – a fact which often creates a gap between the data producers and the data integrators and the informatics world. A good data infrastructure should therefore not place any additional burden on the data-providing scientist (Hardisty & Roberts, 2013; Smith, 2009). On the contrary, if data management and data publication are to become an integral part of normal scientific work, users need to be provided with easy-to-use tools and software that perform all technical aspects of data management and publication, and at the same time offer more benefits than drawbacks (Schnase 2000, Smith 2009) by facilitating and accelerating research and the publication of results.

The Scratchpads software used in this thesis fulfils many of these requirements. It offers

1 http://europa.eu/rapid/press-release_IP-12-790_en.htm, last accessed on 2014-01-20

2 e.g. the Reproducibility Initiative (<https://www.scienceexchange.com/reproducibility/>), see also the special feature by Nature on the topic (<http://www.nature.com/nature/focus/reproducibility/>), both last accessed on 2014-01-20

3 e.g. figshare (<http://www.figshare.com>), DataDryad (<http://www.datadryad.org>), or GigaDB (<http://www.gigadb.org>), all last accessed on 2014-01-20

an easy-to-use web interface which does not require any technical knowledge from the user, and tutorials and an extensive documentation are provided to assist the user during the initial learning phase. Data can be imported into the system with a few clicks and then can be conveniently browsed, searched, annotated and made public (Smith et al. 2011). The software is also ideal for collaborative working. All data imported into the system are automatically formatted as Darwin Core, and thus content can be exposed to large aggregators such as EOL or GBIF, allowing the integration of the data into these systems (Baker et al. 2014). This pre-formatted, standard-conform data structure allows the data to be reused in different contexts – such as the workflow from data assembly and curation to the publication of a manuscript describing a new taxon, as demonstrated in Chapter 2. All data (literature, images, character matrices, specimen data, taxon descriptions) were organised through the system during research phase and the integrated *Publishing Module* then allowed for a seamless creation of the manuscript by simply assembling the different pieces of information and adding additional sections (e.g. introduction, discussion). Through this process, the data used and collected during the research phase is not lost, but is available publicly online, in a format which can be readily integrated into other biodiversity databases or aggregators. Since the writing of Chapter 2, this workflow has been refined during the activities of the European Project ViBRANT⁴ and connected to the recently launched Biodiversity Data Journal⁵, which adopts the philosophy of formatting information as Darwin Core to facilitate reuse of data, the semi-automated assembly of manuscripts and the publication of e.g. distribution datasets through a formal, citable publication (Smith et al. 2013).

Currently, the Scratchpads focus on certain types of data: mainly taxonomic (nomenclatural), bibliographic, specimen data, media (images, video and audio files), species descriptions and occurrence data. Other data types can be integrated either through custom data types or by importing or displaying data from external sources (e.g. articles from Google Scholar or BHL can be displayed). Several such interfaces are already available for simple activation. The modular structure of the software makes it furthermore easy for a person with programming knowledge to create new interfaces to other data types, such as demonstrated in this thesis in Chapter 4: the trait information from the *polytraits* database is displayed on-the-fly within the Scratchpads species pages. This is not a native integration (e.g. in Darwin

4 Virtual Biodiversity Research and Access Network for Taxonomy, <http://www.vbrant.eu>, last accessed on 2014-01-20

5 <http://www.biodiversitydatajournal.com>, last accessed on 2014-01-20

Core Archive export of the information the traits data are not included), but it shows the flexibility of the system.

Unfortunately, it is currently not possible to interactively display the micro-CT data produced in Chapter 3 of this thesis through the system. Current technology is not yet ready to interactively display high-resolution three-dimensional volumetric data through a web browser (see Chapter 3 for details). Developments such as Voluminous⁶ are ongoing, but their integration with the Scratchpads system will need time. Therefore, only preview videos have been stored in the system, which provide a link to the full volumetric dataset for download and manipulation on a local computer.

Tools such as Scratchpads are a good step towards providing incentives for the integration of data publication into the normal scientific workflow. The system is easy to learn and to use and the user does not need to care about data formats, standards or the technical details of data sharing. However, for the system to become even more widely used, a range of improvements could be envisioned, such as integration with other popular tools: online spreadsheets such as Google Docs⁷, a file management system which provides an integration with the local drive of the user (similar to Dropbox⁸), and integrated advanced mapping tools for distribution data. Furthermore, the process of automated publication of data to large aggregators (EOL, GBIF) needs to be facilitated and better advertised to prevent the creation of isolated data silos and to facilitate true linking and integration of data (see also paragraph “From public data to a truly linked network” below).

Another recent initiative incentivising the publication of data is the Biodiversity Data Journal. This journal focuses on publishing datasets and their thorough descriptions. The resulting manuscripts are then citable as normal peer-reviewed articles. At the same time, the software used by the journal ensures that all data are described and published according to community standards such as Darwin Core. Data are directly embedded and marked-up within the document, thus allowing reuse and harvesting by third parties. Such data publishing journals not only provide authors with an incentive to publish their datasets and receive academic credit for them, but the dataset itself is subjected to peer review which ensures a higher quality (Costello et al. 2013a, Page 2013⁹). The *polytraits* database presented in Chapter 4 has been published as a data paper – the first of its kind to describe traits information. The

6 Voluminous is still under development but already in a pre-release version according to the programmer

7 <http://drive.google.com>, last accessed on 2014-01-20

8 <http://www.dropbox.com>, last accessed on 2014-01-20

9 <http://iphylo.blogspot.gr/2013/04/a-decadal-view-of-biodiversity.html>, last accessed on 2014-01-12

paper provides a detailed documentation of data, the availability of metadata, compliance with standards and, of course, the appropriate publication of the dataset. The resulting manuscript is not only fully citable but also acts as a “manual” to understanding, accessing and reusing the data, ensuring a much wider dissemination compared to simply depositing the data in a general online repository. The latter is nowadays required by a number of ecological journals¹⁰, but in contrast to dedicated data papers the simple deposition of data in a repository does not necessarily ensure that the data are in a suitable format for further processing; often there are no requirements for extensive metadata documentation, and peer review of the data is not guaranteed.

Tools such as Scratchpads and data journals can provide a means to promote data publication since the real incentive for the user is not the data publication process. By being “rewarded” by the end product (either a citable publication or a software making daily research more structured and easy), researchers are more likely to adopt good practices without experiencing the data management process as a burden to their daily work schedule.

A need for new data types

The field of biodiversity informatics has achieved remarkable developments concerning the availability and integration of biodiversity information. Initiatives such as OBIS, GBIF, various taxonomic registers, or the Biodiversity Heritage Library provide vast volumes of data and information which have revolutionised the every-day work of biodiversity scientists. The integration of data from distributed sources facilitates large-scale analyses and new discoveries arising from recombinations of data (e.g. Arvanitidis et al. 2009, Somerfield et al. 2009, Tittensor et al. 2010, Webb et al. 2010, Costello et al. 2013b). Most of the effort in the field however has so far been directed towards the digitisation and integration of taxonomic (nomenclature, classifications, specimens) and species occurrence data. The popularity of Darwin Core as a common standard for these taxon-centric data and a large body of guidelines and best practice manuals may have promoted the increasing availability of specimen and occurrence data, for which the standard was developed.

In order to really start understanding processes in our biosphere and to model the interactions between organisms and their environment, distribution data alone are not

¹⁰ e.g. <http://esapubs.org/esapubs/DataReg.htm>,
http://www.britishecologicalsociety.org/wp-content/uploads/Data_Archiving_QA_2014.pdf or
<http://www.nature.com/authors/policies/availability.html>, all last accessed on 2014-01-20

sufficient. New types of information are needed to describe these processes (Hardisty et al. 2013, Purves et al. 2013). In order to assess phenomena such as climate change, ocean acidification, biodiversity loss and other human-induced changes to the environment, data about the biological characteristics of species are needed, covering aspects of morphology, reproduction, environmental preferences, vulnerability, interactions with other species and the role of the species in the ecosystem (Hampton et al. 2013, Harfoot and Roberts 2014). At the local scale, this type of data is essential in order to answer “what if” questions: environmental managers and policy makers need to be able to assess potential changes in ecosystem functioning as a result of the loss or introduction of species. Such questions cannot be answered with species occurrence data alone. The approach taken in Chapter 5 of this thesis incorporates, therefore, two complementary types of information: species attribute data in combination with species occurrence data. The predictive models developed during this study thus present a powerful way of assessing the potential loss of species and identifying the functionally important components of the ecosystem, allowing decision makers to take appropriate conservation measures. Projected on a global scale, such predictive models incorporating trait data allow for novel insights into the functioning of our biosphere and potential responses of ecosystem functioning to disturbances (e.g. Diaz and Cabido 1997, Le Quéré et al. 2005, Díaz et al. 2006, Bonada et al. 2007, Scheiter et al. 2013).

To date, the available data on species traits are, however, still scarce compared to species occurrence data. Traits information is rarely organised and published in a structured, standardised way. Most data, if published at all, are provided as supplementary material to publications, in formats which are not computer-readable. Few traits databases exist (e.g. BIOTIC, TRY, NMiTA, Sea Life Base) but they do not follow a standardised way of storing and describing these data. This results in a gappy, fragmented landscape of trait information, reminiscent of the early situation concerning occurrence data. The lack of standardised data formats for annotating trait data and the absence of dedicated repositories provide an obstacle to the massive public availability of this type of information. Requirements for giving a boost to an increased publication and availability of trait information are:

- a) A widely accepted community standard to describe the data (Darwin Core in its current form is a poor solution to convey trait information). Ideally, the community needs to develop cross-taxon and cross-purpose ontologies which standardise trait concepts and define the relationships between these concepts. Ontologies are a necessary prerequisite to integrate

trait data into the semantic web and will make the information fully exploitable by both humans and computers. Such a standardisation prevents isolation and incompatibility of information. A first attempt to this has been taken in Chapter 4 of this thesis: information has been mapped to existing ontology concepts (where these existed), facilitating the direct integration of the *polytraits* data with other data in EOL's TraitBank.

- b) Mass digitising of legacy literature and text-mining tools that can extract relevant information from scientific literature (Parr et al. 2011).
- c) Public repositories which act as integrators, promote good practice and prevent the fragmentation of information. An attempt to this is EOL's recently released TraitBank, which could become an important aggregator for this type of data. EOL supports the semantic annotation of trait data, facilitating integrated, intelligent queries across data.

Another area which has so far been neglected by digitisation efforts are museum collections, possibly the largest and most important authoritative source of biodiversity data (Balke et al. 2013). While specimen metadata were some of the first data to be available in digital forms since museums were applying electronic collection management software, the actual specimens are rarely available in a digital format. The availability of reference material – type specimens – is however one of the most important pillars of taxonomic research. The tedious, time-consuming process of obtaining physical specimens for study is one of the major bottlenecks in taxonomic research. In contrast to other disciplines, alpha taxonomy in its current form profits only little from the rapid technological developments of the last decades. The production of new data is still extremely slow, manual and results in “analogue” data: simple texts as well as pictures and illustrations, which are not exploitable by computers (although this is changing for taxon descriptions, see Chapter 2 of this thesis). Thus, taxonomy in its current form does not scale and is not compatible with the pace other disciplines (such as the digital-born *-omics* research areas) are setting. Morphology-based conventional taxonomy needs to play in the same league as these data-centric disciplines in order to survive. Chapter 3 of this thesis presents an innovative approach to help taxonomy to overcome its major problem: the immediate access to type material. The method of creating three-dimensional morphological data and the application of the cybertype concept presented in this study could in the future revolutionise taxonomic work. The instant and simultaneous availability of high-resolution virtual type material could not only accelerate taxonomic revisions but form

the basis for a completely new branch of phylogenetic research, based on quantitative, digitally-born morphology data, thus turning taxonomy into a real cyber-discipline.

Much technological process is, however, still needed to accomplish the automated mass digitisation of data (Balke et al. 2013), and the complexity of digital, three-dimensional morphological data – both in content as well as in volume (often several gigabytes per dataset) – require a specific technical infrastructure and dedicated approaches to data management, archival and dissemination. Chapter 3 provides an extensive roadmap for future research required by the biodiversity informatics and museum communities in order to realise virtual collections. Similar to traits data, the current biodiversity informatics infrastructure is not yet ready to accommodate massive amounts of virtual specimen datasets. The most urgent priorities identified in this study are, therefore, the development of large public repositories and community standards needed to describe, retrieve and integrate these data. The biodiversity informatics community has immediately taken up these challenges, and since the publication of the manuscript in Chapter 3 some progress has been made towards these directions: First attempts to a standardised, integrated to a standardised, integrated publication of micro-CT datasets including descriptive metadata have been attempted using both GigaDB and the ISA-TAB standard (Sansone et al. 2012, Edmunds et al. 2013, Stoev et al. 2013), but more work is required to ensure that three-dimensional data do not develop into isolated data worlds but that community standards are developed to describe the datasets in a way that they can be retrieved, cross-linked and integrated. Again, dedicated repositories with intuitive interfaces (e.g. preview of the dataset, interactive manipulation) will support this process. Virtual biodiversity collections might be a vision for the future, but the first museums have recognised the potential already and are starting to digitise their collections (Smith and Blagoderov 2012, Ang et al. 2013a, Balke et al. 2013), and despite the present-day difficulties, virtual collections will eventually form a central part of the cybertaxonomy landscape.

Finally, the creation of three-dimensional virtual specimens will not only revolutionise the availability of type material, but also taxonomic publishing. Currently, despite the available technology, the two-dimensionality of taxonomic publications has continued into the digital age: even electronic forms of articles still contain two-dimensional representation of three-dimensional structures. Chapter 3 of this thesis gives an outlook to the potential future of taxonomic publishing by directly embedding three-dimensional models of specimens into the taxonomic literature, providing a holistic view on the morphological features of the taxon (Ang

et al. 2013b, Stoev et al. 2013).

From public data to a truly linked network

Efforts to make more data from various thematic fields digitally available are continuously ongoing (Thessen and Patterson 2011, Enke et al. 2012, Costello et al. 2013a, Smith et al. 2013), yet data are published across a number of scattered databases, files, repositories and in journal articles, often in formats that are not exploitable by computers, and without proper documentation through metadata. This growing amount of digital biodiversity information has drawn attention to the fact that new mechanisms are needed to manage and integrate these data across an increasing number of sources and formats in order to make best use of them (Catapano et al. 2011, Parsons et al. 2011, Minami et al. 2013). The eventual goal is to make the data ready for integrative, large-scale analyses and global ecosystem models, but since such processes require computing power, we therefore need to ensure that the existing data are available in a format which can be used by computers to compare them, detect patterns and infer new information (Schnase 2000, Page 2006, 2008, Smith 2009). Community standards ensuring the interoperability of data are increasingly being used, and data are being transformed into machine-readable formats, but the current technological infrastructure still does not allow for a seamless integration of these data across distributed sources. The great challenge for the biodiversity informatics community is therefore to create a network of Linked Open Data (LOD) which allows for computer-mediated, complex queries across sources and disciplines, in short: a distributed knowledgebase for biodiversity (Parr et al. 2011, Thessen et al. 2011).

Among the most pressing requirements towards the creation of this knowledgebase are:

- a) The creation and consistent use of permanent, resolvable, globally unique identifiers (GUIDs) to be able to unambiguously refer to pieces of information, data objects and concepts. The lack of GUIDs is a major gap in the current data landscape which has been repeatedly been pointed out (Page 2009, Parr et al. 2011, Thessen et al. 2011, Hardisty et al. 2013). These identifiers need to be stable and resolvable in order to maintain the integrity of the network. One of the first – and most pressing – areas where GUIDs should be consistently used are taxon names. Biodiversity is a taxon-centric science, therefore most information is arranged around the elements in the tree of life – taxa or taxon concepts. Taxonomic names and their classification have often been proposed to serve as globally unique identifiers for taxa, even as ontologies (Patterson et al. 2008, 2010, Minami et al.

2013, Tuominen et al. 2013), but this approach has also been criticised, since taxonomic names do not fulfil all requirements for such identifiers: they are not necessarily unique (homonyms exist), and their dynamic nature (continuous changes through taxonomic revisions) would make such a system unstable (Page 2008). Furthermore, the same taxonomic name can denote different concepts (“*sensu* Author X”), another aspect which makes taxonomic names weak identifiers in a semantic web. However, the initial approach of using taxa to link information is still valid. In a linked data network, we need to de-couple taxon names from their meaning by assigning globally unique, shared identifiers to taxonomic concepts and define their interrelationships. The development of such a taxonomic backbone of identifiers is one of the major prerequisites for linking biodiversity information (Hardisty et al. 2013).

- b) The creation and use of ontologies and vocabularies for specific domains of biodiversity for describing content. Currently, Darwin Core is widely used as a standard, but operates on a metadata level, describing the type of content, not the content itself, or as stated by Page (2013)¹¹: “*The most successful vocabulary we’ve come up with (Darwin Core) is essentially an agreed way to label columns in Excel spreadsheets*”. In combination with controlled vocabularies, the content described by Darwin Core terms would become much more exploitable, since the same content across different data sources would be described by the same terms/ identifiers, ensuring comparability. Semantic approaches need to be introduced and applied to enable machine logic across distributed sources. Most ontologies currently in use describe domains in the molecular sciences such as genetics, genomics, transcriptomics, etc. (Parr et al. 2011).
- c) The promotion and increased use of computer-readable licenses for data publication that allow the reuse and recombination of existing data, such as Creative Commons¹². Currently, existing datasets are described by a plethora of custom-made or human-only readable licenses which often prevent the data from being used in any other way than simple display¹³. Such datasets often cannot be integrated into a network of linked data due to legal restrictions. The wider application of computer-readable licenses is often simply a matter of increased promotion and education of scientists.
- d) The development of user-friendly applications that demonstrate the added value of such a

11 <http://iphylo.blogspot.gr/2013/11/zookeys-gbif-and-github-fixing-darwin.html>, last accessed on 2014-01-12

12 <http://www.creativecommons.org>, last accessed on 2014-01-22

13 See blog post by Peter Desmet: <http://peterdesmet.com/posts/analyzing-gbif-data-licenses.html>, last accessed on 2014-01-22

linked network to the non-technological user. We need “proof of concept” tools to encourage data sharing and the increased use of standards, ontologies and GUIDs. An example for such an application has been presented by Bolleman et al. (2012) who use a semantic infrastructure to discover inconsistencies in datasets. Semantics could therefore be used to improve data quality of existing datasets such as GBIF, and such a tool would certainly prove the usefulness of the approach.

Several of the above mentioned elements are already being implemented or actively used (see e.g. Bolleman et al. 2012; Box 3 in Parr et al. 2011; or the Biological Science Collections project¹⁴) and progressively, more initiatives provide marked-up data, DOIs and Life Science Identifiers (LSIDs)), but coordinated approaches are needed to really make the existing data interoperable on a large scale.

In this thesis, elements of linked and standardised data can be found throughout all chapters. The ms in Chapter 2 has been published as a marked-up version with machine-readable links to information in other repositories such as Zoobank, Plazi and EOL, in Chapters 3 and 5 where DOIs ensure the retrievability of the underlying data in Dryad and figshare, and in Chapter 4, which makes use of globally unique identifiers in the form of URIs, linking back to ontology concepts. The latter allows not only the integration of the data with existing data in the Encyclopedia of Life's TraitBank, but the use of ontologies also opens up possibilities for semantic queries. All data in this thesis have been released under an appropriate, machine-readable Creative Commons License (CC-BY) which allows free reuse of the data. Other elements of this thesis make extensive use of Darwin Core, allowing integration and reuse of data elements; however, these elements are not truly linkable since they do not always carry GUIDs. The Scratchpads system does provide a DWC-A export of the data in the Scratchpads, but embedded RDF into HTML pages to make the web-pages computer-readable is not yet fully supported (Smith et al. 2011). Drupal natively supports the creation of RDF (Thompson and Richard 2013) but the Scratchpads do not yet support APIs or embedded RDF out-of-the-box. A more automated integration with the rest of the biodiversity data cloud is therefore needed for the future in order to prevent the creation of isolated data. The latter has been recognised by the development team and is being addressed (Smith et al. 2011). A number of additional data objects created during this thesis hold potential to be described and stored in a more adequate way, by introducing resolvable GUIDs and annotating certain objects with

¹⁴ <http://biscicol.blogspot.com/>, last accessed: 2014-01-22

more metadata so that they are not “lost” – these issues will be addressed during ongoing data curation.

Conclusions

Biodiversity research is currently undergoing a silent revolution. Digital data are being created in large volumes and at a rapid pace and new types of information are emerging. Existing information has begun to be linked together, whereas intelligent algorithms can help in transforming this ocean of data into an intelligent network of knowledge. Yet, without clear rules and standards for these new data types it is not possible to exploit the true potential of this wealth of information. A clear focus in future biodiversity informatics research needs to lie on linking the distributed bits and pieces together. Data need to be documented and standardised to be compatible with existing information. Elements on the way to this semantic “database of everything” (Smith 2009) will include permanent, globally unique identifiers for the items of this network, especially for taxon names or taxon concepts, as well as adding semantics to the data so that they can be understood by computers, which, in turn, will allow scientists to ask complex questions and find these answers in the data. Such an infrastructure will allow unforeseen and powerful analyses and discoveries, creating a revolution in our abilities to understand, model and predict the biological processes on our planet. However, at the basis of the resolution still lies the every-day scientific practice of each single researcher in the wider community. Sociological barriers to the creation of such a linked data network are greater than technical barriers, and data management and data sharing are still not an integrated part of scientific practice. Within the overarching framework of this thesis, the papers presented here demonstrate an approach to modern biodiversity research by using existing and newly created tools and by linking, sharing and mobilising data, always with the aim of making the underlying data and results reusable in a wider context. The published papers set standards in their respective fields and have already inspired additional research:

- a) The first chapter was one of the first taxonomic papers ever to be published through an integrated online writing tool in a virtual research environment. This workflow has now become a normal part of the Biodiversity Data Journal, implementing a revolutionary way of publishing taxonomic and other data.
- b) The second chapter introduces the concept of three-dimensional cybertypes and identifies important research directions for the biodiversity informatics and museum communities

towards the creation of virtual galleries and the curation of cybertypes. This paper has inspired the first ever taxonomic description of a new species including the creation of a cybertype (Stoev et al. 2013).

- c) Chapter 4 is the first data paper to describe a semantically annotated dataset of biological trait information and to make use of EOL's TraitBank as a repository for the data.

Furthermore, by using a dedicated online content management, all research data underpinning this thesis can be accessed through a central website and are freely available for reuse.

All the above finally share a common feature: the main scope of each study was the testing of a specific biological hypothesis, facilitated by the progress made in the field of biodiversity informatics and cybertaxonomy and creating new synergies. Work practices such as the ones presented in this thesis should, therefore, be adopted by every researcher in the biodiversity community. Biodiversity data are too valuable to be lost, and in the combined effort of protecting biodiversity every small piece of information helps in understanding the big picture of life on earth.

References

- Ang Y, Puniamoorthy J, Pont AC, Bartak M, Blanckenhorn WU, Eberhard WG, Puniamoorthy N, Silva VC, Munari L, Meier R (2013) A plea for digital reference collections and other science-based digitization initiatives in taxonomy: Sepsidnet as exemplar. *Systematic Entomology* 38: 637–644. doi:10.1111/syen.12015
- Ang Y, Wong LJ, Meier R (2013) Using seemingly unnecessary illustrations to improve the diagnostic usefulness of descriptions in taxonomy—a case study on *Perochaeta orientalis* (Diptera, Sepsidae). *ZooKeys* 355: 9–27. doi:10.3897/zookeys.355.6013
- Arvanitidis C, Somerfield PJ, Rumohr H, Faulwetter S, Valavanis VD, *et al.* (2009) Biological geography of the European seas: results from the MacroBen database. *Marine Ecology Progress Series* 382: 265–278. doi:10.3354/meps07955
- Baker E, Rycroft S, Smith VS (2014) Linking multiple biodiversity informatics platforms with Darwin Core Archives. *Biodiversity Data Journal* 2: e1039. doi:10.3897/BDJ.2.e1039
- Balke M, Schmidt S, Hausmann A, Toussaint E, Bergsten J, Buffington M, Häuser CL, *et al.* (2013) Biodiversity into your hands - A call for a virtual global natural history 'metacollection'. *Frontiers in Zoology* 10: 55. doi:10.1186/1742-9994-10-55
- Bizer C, Jentsch A, Cyganiak R (2011) State of the LOD cloud. Available from: http://wiki.planet-data.eu/web/State_of_the_LOD_cloud.
- Bolleman J, Gehant S, The Uniprot Consortium (2012) Catching inconsistencies with the semantic web: a biocuration case study. In: SWAT4LS 2012: Paris, France. pp. 1–12.

- Bonada N, Dolédec S, Statzner B (2007) Taxonomic and biological trait differences of stream macroinvertebrate communities between mediterranean and temperate regions: implications for future climatic scenarios. *Global Change Biology* 13: 1658–1671. doi:10.1111/j.1365-2486.2007.01375.x
- Catapano T, Hobern D, Lapp H, Morris RA, Morrison N, Noy N, Schildhauer M, Thau D (2011) Recommendations for the Use of Knowledge Organisation Systems by GBIF. Copenhagen, 49 pp. Available from: http://links.gbif.org/gbif_kos_whitepaper_v1.pdf.
- Chavan VS, Ingwersen P (2009) Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. *BMC Bioinformatics* 10 Suppl 1: S2. doi:10.1186/1471-2105-10-S14-S2
- Costello MJ (2009) Motivating Online Publication of Data. *BioScience* 59: 418–427. doi:10.1525/bio.2009.59.5.9
- Costello MJ, Bouchet P, Boxshall G, Fauchald K, Gordon D, Hoeksema BW, Poore GCB, van Soest RWM, Stöhr S, Walter TC, Vanhoorne B, Decock W, Appeltans W (2013) Global Coordination and Standardisation in Marine Biodiversity through the World Register of Marine Species (WoRMS) and Related Databases Browman H (Ed). *PLoS ONE* 8: e51629. doi:10.1371/journal.pone.0051629
- Costello MJ, Michener WK, Gahegan M, Zhang Z-Q, Bourne PE (2013) Biodiversity data should be published, cited, and peer reviewed. *Trends in ecology & evolution* 28: 454–61. doi:10.1016/j.tree.2013.05.002
- Diaz S, Cabido M (1997) Plant functional types and ecosystem function in relation to global change. *Journal of Vegetation Science* 8: 463–474. doi:10.2307/3237198
- Díaz S, Lavorel S, McIntyre S, Falczuk V, Casanoves F, Milchunas DG, Skarpe C, Rusch G, Sternberg M, Noy-Meir I, Landsberg J, Zhang W, Clark H, Campbell BD (2006) Plant trait responses to grazing - a global synthesis. *Global Change Biology* 12: 1–29. doi:10.1111/j.1365-2486.2006.01288.x
- Edmunds SC, Hunter CI, Smith VS, Stoev P, Penev L (2013) Biodiversity research in the “big data” era: GigaScience and Pensoft work together to publish the most data-rich species description. *GigaScience* 2: 14. doi:10.1186/2047-217X-2-14
- Enke N, Thessen A, Bach K, Bendix J, Seeger B, Gemeinholzer B (2012) The user’s view on biodiversity data sharing – Investigating facts of acceptance and requirements to realize a sustainable use of research data –. *Ecological Informatics* 11: 25–33. doi:10.1016/j.ecoinf.2012.03.004
- Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, Duke CS, Porter JH (2013) Big data and the future of ecology. *Frontiers in Ecology and the Environment*: 130312142848005. doi:10.1890/120103
- Hardisty A, Roberts D, Addink W, Aelterman B, Agosti D, *et al.* (2013) A decadal view of biodiversity informatics: challenges and priorities. *BMC ecology* 13: 16. doi:10.1186/1472-6785-13-16
- Harfoot M, Roberts D (2014) Taxonomy: Call for ecosystem modelling data. *Nature* 505: 160. doi:10.1038/505160a
- Huang X, Qiao G (2011) Biodiversity databases should gain support from journals. *Trends in*

- ecology & evolution 26: 1632–1633. doi:10.1016/j.tree.2011.05.006
- Minami Y, Takeda H, Kato F, Ohmukai I (2013) Towards a Data Hub for Biodiversity with LOD. In: Proceedings of the Second Joint International Conference, JIST 2012, Nara, Japan, December 2-4, 2012. , 356–361. doi:10.1007/978-3-642-37996-3_29
- Page RDM (2008) Biodiversity informatics: the challenge of linking data and the role of shared identifiers. *Briefings in Bioinformatics* 9: 345–354.
- Parr C, Guralnick R, Cellinese N, Page RDM (2011) Evolutionary informatics: unifying knowledge about the diversity of life. *Trends in ecology & evolution* 27: 94–103. doi:10.1016/j.tree.2011.11.001
- Parsons M a., Godoy O, LeDrew E, de Bruin TF, Danis B, Tomlinson S, Carlson D (2011) A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science* 37: 555–569. doi:10.1177/0165551511412705
- Purves D, Scharlemann J, Harfoot M, Newbold T, Tittensor DP, Hutton J, Emmot S (2013) Time to model all life on Earth. *Nature* 493: 7–9.
- Le Quéré C, Harrison SP, Prentice IC, Buitenhuis ET, Aumont O, *et al.* (2005) Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology* 11: 2016–2040. doi:10.1111/j.1365-2486.2005.01004.x
- Reichman OJ, Jones MB, Schildhauer MP (2011) Challenges and Opportunities of Open Data in Ecology. *Science* 331: 703–705. doi:10.1126/science.1197962
- Sansone S-A, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, *et al.* (2012) Toward interoperable bioscience data. *Nature genetics* 44: 121–6. doi:10.1038/ng.1054
- Scheiter S, Langan L, Higgins SI (2013) Next-generation dynamic global vegetation models: learning from community ecology. *New Phytologist*: n/a–n/a. doi:10.1111/nph.12210
- Schnase J (2000) Research directions in biodiversity informatics. In: Proceedings of the 26th International Conference on Very Large Databases, Cairo, Egypt, 2000. Cairo, 697–700.
- Smith VS (2009) Data publication: towards a database of everything. *BMC research notes* 2: 113. doi:10.1186/1756-0500-2-113
- Smith VS, Blagoderov V (2012) Bringing collections out of the dark. *ZooKeys* 209: 1–6. doi:10.3897/zookeys.209.3699
- Smith VS, Georgiev T, Stoev P, Biserkov J, Miller J, Livermore L, Baker E, Mietchen D, Couvreur TLP, Mueller G, Dikow T, Helgen KM, Frank J, Agosti D, Roberts D, Penev L (2013) Beyond dead trees: integrating the scientific process in the Biodiversity Data Journal. *Biodiversity Data Journal* 1: e995. doi:10.3897/BDJ.1.e995
- Smith VS, Rycroft SD, Brake I, Scott B, Baker E, Livermore L, Blagoderov V, Roberts D (2011) Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science. *ZooKeys* 150: 53–70. doi:10.3897/zookeys.150.2193
- Somerfield PJ, Arvanitidis C, Vanden Berghe E (2009) Large-scale studies of the European benthos: the MacroBen database. *Marine Ecology Progress Series* 382: 221–224. doi:10.3354/meps08045
- Stodden V, Guo P, Ma Z (2013) Toward Reproducible Computational Research: An Empirical

- Analysis of Data and Code Policy Adoption by Journals. PLoS one 8: e67111.
doi:10.1371/journal.pone.0067111
- Stoev P, Komericki A, Akkari N, Liu S, Zhou X, Weigand AM, Hostens J, Hunter CI, Edmunds SC, Porco D, Zapparoli M, Georgiev T, Mietchen D, Roberts D, Faulwetter S, Smith VS, Penev L (2013) *Eupolybothrus cavernicolus* Komericki & Stoev sp. n. (Chilopoda: Lithobiomorpha: Lithobiidae): the first eukaryotic species description combining transcriptomic, DNA barcoding and micro-CT imaging data. Biodiversity Data Journal 1: e1013. doi:10.3897/BDJ.1.e1013
- Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M (2011) Data sharing by scientists: practices and perceptions. PLoS ONE 6: e21101. doi:10.1371/journal.pone.0021101
- Thessen AE, Patterson D (2011) Data issues in the life sciences. ZooKeys 150: 15. doi:10.3897/zookeys.150.1766
- Thompson K, Richard J (2013) Moving Our Data to the Semantic Web: Leveraging a Content Management System to Create the Linked Open Library. Journal of Library Metadata 13: 290–309. doi:10.1080/19386389.2013.828551
- Tittensor DP, Mora C, Jetz W, Lotze HK, Ricard D, Berghe E Vanden, Worm B (2010) Global patterns and predictors of marine biodiversity across taxa. Nature 466: 1098–1101. doi:10.1038/nature09329
- Webb TJ, Vanden Berghe E, O'Dor R (2010) Biodiversity's Big Wet Secret: The Global Distribution of Marine Biological Records Reveals Chronic Under-Exploration of the Deep Pelagic Ocean. PLoS ONE 5: e10223. doi:10.1371/journal.pone.0010223

Glossary of terms

Glossary of terms

acicula	A stout bristle supporting the present in the <u>parapodia</u> of certain <u>polychaete</u> taxa
acuminate	Tapering to a sharp point
aggregator	Large initiatives such as <u>EOL</u> or <u>GBIF</u> which do not create data but assemble, integrate and disseminate data from a variety of sources
ALA	Atlas of Living Australia (http://www.ala.org.au): a species information system on the Australian Biodiversity.
Algaebase	A global database of information on algae (http://www.algaebase.org)
alpha taxonomy	<u>Taxonomy</u> in its classical sense: finding, describing and naming taxa
anaglyph 3D glasses	Glasses where one eye has a red filter, the other a cyan/blue/green one, used to create a 3D-effect when viewing anaglyph images (images that contain two differently filtered coloured images, one for each eye).
API	see <u>application programming interface</u>
application programming interface	Specification of how different parts of software can interact with each other
AVI	Audio Video Interleave: proprietary video format by Microsoft
BDJ	Biodiversity Data Journal (http://www.biodiversitydatajournal.com): an <u>open access</u> scientific journal focusing on the publication of biodiversity-related data of any kind
BEF	Biodiversity and Ecosystem Functioning
BHL	Biodiversity Heritage Library (http://biodiversitylibrary.org), a global project dedicated to the mass digitization of legacy biodiversity literature
biogeographic data	Data on the distribution of species, in biodiversity informatics often synonymous with occurrence data in the form of (at least): taxon name, latitude, longitude and depth / elevation
BIOTIC	Biological Traits Information Catalogue (http://www.marlin.ac.uk/biotic): a database of biological traits of marine organisms with a focus on species around the British Isles
bmp	Bitmap: an image format
boolean	A data type used in programming languages which can only have two values, <i>true</i> or <i>false</i>
BTA	Biological Traits Analysis: an approach of assessing ecosystem functioning by using information on species' life cycle traits
C++	A programming language
caruncle	A large sensory organ on the <u>prostomium</u> of <u>polychaetes</u> in the clade Amphinomida
CC-BY	<u>Creative Commons</u> Attribution License: a copyright license that allows unlimited reuse of the licensed object, provided credit to the author is given
CCD camera	Charge-coupled device camera: a technology used in digital photography

	to convert a light signal into a digital value
chaetae	Stiff, chitinous bristles in <u>polychaetes</u> , located on the <u>parapodia</u>
chaetiger	Segment in a <u>polychaete</u> worm which bears <u>chaetae</u>
cirrus	Fleshy appendages on the body of a <u>polychaete</u> , often long and thread-like
cLSM	Confocal laser scanning microscopy: a technique for obtaining three-dimensional high-resolution images
CMS	see <u>Content Management System</u>
COL	Catalogue of Life (http://www.catalogueoflife.org): a global database attempting to catalogue all known species
compound chaeta	<u>Chaeta</u> composed of a shaft and a blade, divided by a joint
content management system	A (mostly online) software that allows collaborative publishing, editing and modifying content through a central interface
Creative Commons licenses	A set of copyright-licenses that describe which rights of the licensed material are reserved and to what extent and under which condition the material can be modified and shared. The licenses are standardised, simple to understand and <u>computer-readable</u> (http://www.creativecommons.org).
cryptic species	Morphologically indistinguishable species that differ only in their molecular or life cycle characteristics
csv	Comma /character separated value: a plain text format in which individual fields are separated by a specified character
CT	Computer tomography, a non-destructive imaging technique that uses a set of X-ray images to calculate three-dimensional representations of the imaged object
CTVox	A commercial <u>volume rendering</u> software developed by SkyScan / Bruker
Darwin Core	A standardised vocabulary of terms used in biodiversity informatics to facilitate the discovery, retrieval, and integration of organism-related information
data object	A structured representation of data. This includes the object itself and its <u>metadata</u> .
data paper	A novel form of peer-reviewed scientific publication which accompanies the publication of a dataset by describing its scope, spatio-temporal, taxonomic and other content as well as involved persons and usage rights. Initiated to incentivise the publication of datasets by allowing authors to receive academic credit for their efforts through citation.
DICOM	Digital Imaging and Communications in Medicine: a standard providing format definitions and a communication protocol for the description and exchange of medical image data
distal	The point of an anatomical appendage which is furthest from the attachment to the main body
DOI	Digital Object Identifier: a digital identifier used to uniquely identify an object and retrieve metadata about it. Mostly applied to documents such as scientific papers, books, reports, etc. but not restricted to these – any

	digital object can be assigned a DOI
dorsal cirrus	A <u>cirrus</u> situated on the upper part of the <u>notopodium</u>
Drishti	Free <u>volume rendering</u> software (http://sf.anu.edu.au/Vizlab/drishti)
Dropbox	An online file storage service which integrates with the users computer by creating a special folder: all files saved in that folder are uploaded to the user's online storage space.
Drupal	An open-source and free online content management system (http://drupal.org)
Dryad / Datadryad	An online data repository to store data underlying scientific publications. Assigns DOIs to the submitted data so that they can be discovered and cited (http://datadryad.org).
DwC	<u>Darwin Core</u>
DwC-A	Darwin Core Archive: a set of files employing the <u>Darwin Core</u> standard, arranged around a file called meta.xml which describes the relationship between the individual files. Recommended format to publish data within the <u>GBIF</u> network.
EOL	Encyclopedia of Life (http://www.eol.org): an international initiative aggregating and disseminating data and scientific knowledge about all known species.
falciger	Compound <u>chaeta</u> with a stout / blunt / hooked tip
fascicle	Bundle of <u>chaetae</u>
figshare	An online data repository to store data underlying scientific publications. Assigns DOIs to the submitted data so that they can be discovered and cited (http://www.figshare.com).
Fishbase	A global species information system on fish (http://www.fishbase.org)
FLV	Flash Video: video format used by Adobe's Flash Player
GB	Gigabyte: 10 ⁹ bytes. A byte is a unit of digital information, its multiples (kilo-, mega-, gigabyte) are often used to describe the size of a digital storage medium, file or dataset.
GBIF	Global Biodiversity Information Facility (http://www.gbif.org): a global data infrastructure, providing free access to biodiversity data, developing and promoting standards and tools for data management and publication
GenBank	A global database of annotated nucleotide sequences (http://www.ncbi.nlm.nih.gov/genbank)
GigaDB	An online data repository to store data underlying scientific publications, specializing on very large datasets. Assigns <u>DOI</u> s to the submitted data so that they can be discovered and cited (http://www.gigadb.org).
GNA	Global Names Architecture (http://www.globalnames.org): an initiative trying to build an infrastructure based on scientific names to index biological content
GUID	Globally Unique Identifier: a globally unique alphanumeric sequence of characters which is used to unambiguously refer to a <u>data object</u>

harvesting	Extraction of data from web sites or databases. Large <u>aggregators</u> such as <u>OBIS</u> , <u>GBIF</u> , etc. harvest data from associated projects and sites in order to integrate these data with others and make them available through a central portal.
HMDS	Hexamethyldisilazane: a chemical used for drying specimens without affecting their morphology
holotype	The physical specimen upon which the description of a new species is based
HTML	Hypertext Markup Language: the main <u>markup</u> language for creating web pages
HTTP	Hypertext Transfer Protocol: the communication standard used to transfer data on the World Wide Web
ICZN	International Commission of Zoological Nomenclature (http://www.iczn.org): an organisation advising on and regulating the correct use of the scientific names of animals
IF	Index Fungorum (http://indexfungorum.org): an international project to index all scientific names of fungi
integer	A numeric datatype in computer programming which is designed to store a whole number only
ION	Index of Organism Names (http://www.organismnames.com): a database of organism names and related data gathered by Thomson Reuters from the scientific literature
IPNI	International Plant Names Index: a database on plant names, nomenclature and associated bibliographies (http://www.ipni.org)
ISA-TAB	A newly proposed data format intended to facilitate integration of diverse datasets across various life-science disciplines (http://www.isa-tools.org)
isosurface model	Geometrical, three-dimensional model consisting of surfaces of equal sizes (often triangles)
jpg	Joint Picture Group. Image format
JSON	JavaScript Object Notation: a data format using human-readable text to transmit <u>data objects</u> in the form "attribute:value"
kV	Kilovolt
limbate chaetae	A type of simple <u>chaeta</u> which widens into a "winged" structure along its distal half
LOD	The totality of freely accessible data on the internet which can be identified and are linked to other data via <u>URIs</u> . Often, these data are also described through machine-readable formats such as <u>RDF</u> and <u>OWL</u> , thus forming a part of the <u>semantic web</u> .
LSID	Life Science Identifiers: A naming standard to uniquely identify shared <u>data objects</u> in life sciences, including biodiversity informatics, in the form <i>urn:lsid:authority:namespace:object:revision</i>
machine-readable format,	Data which are provided in a format that allows computers to interpret them by identifying specific elements, relating them to each other

computer-readable format	according to specified sets of rules and thus being able to potentially infer new information. Examples for machine-readable formats are RDE , OWL , XML , JSON or, to a certain extent, CSV .
mandibles	Ventral jaw plate of the jaw apparatus in the polychaete clade Eunicida
markup, markup language	A computer language that uses tags (“wrappers”) to annotate elements in a document so that these can be identified by computers.
maxillae	Dorsal jaw plate of the jaw apparatus in the polychaete clade Eunicida, consisting of several maxillary pairs
MDS	Multidimensional Scaling: an ordination method used to visualise the level of similarity between pairs of items
metadata	Data about data. Ambiguous term, since data can be both “normal data” as well as “metadata”, depending on the context. Metadata help to understand the context of the data as well as provide a means to integrate it with other data.
MI, MII, MIII, MIV, ...	Enumerations of the different maxillary pairs in the polychaete clade Eunicida
μA	Micro-ampere
micro-CT	Micro-computed tomography: Computer tomography for small samples in high resolution
module	Independent software part which performs a certain function. In the content management system Drupal , modules are pieces of software which can be “plugged” into the main software core in order to extend its functionality.
morphometric analysis	Statistical analysis based on the shape and size of morphological characters of organisms to detect similarities or significant differences between samples, populations or species
MRI	Magnetic Resonance Imaging: a non-destructive, three-dimensional imaging method based on the reaction of atoms to strong magnetic fields
mysql	An open-source relational database management software (http://www.mysql.com)
neotype	A specimen selected to serve as the type specimen when the original holotype was lost, destroyed or never specified.
neurochaeta	Chaeta of a neuropodium .
neuropodium	Ventral part of a parapodium
NMiTA	Neogene Marine Biota of Tropical America (http://eusmilia.geology.uiowa.edu): an online species database containing also traits information. Focuses on species of Tropical America.
notochaeta	Chaeta of a notopodium
notopodium	Dorsal part of a parapodium
OBIS	Ocean Biogeographic Information System (http://iobis.org): a global biogeographic database on marine species
OBJ	Wavefront Object: a data format used to represent three-dimensional geometric objects

-omics	A group of biological disciplines whose names end in -omics, such as genomics, transcriptomics, metabolomics. Often used to describe highly digitised disciplines that produce and analyse extremely large quantities of data.
ontology	A set of concepts and their interrelationships which are described through a standardised vocabulary
open access	Free access to scientific articles and datasets via the internet
Open Data Commons	An initiative similar to Creative Commons , providing licenses for free databases (http://opendatacommons.org)
OPT	Optical Projection Tomography: a non-destructive, three-dimensional, high-resolution imaging method based on optical microscopy
OWL	A computer language/ data format used for knowledge representation, e.g the creation of ontologies
palps	A paired set of cushion-like or tentacular sensory organs at the prostomium of many polychaete taxa
paragnaths	Small denticles on the pharynx of polychaetes in the family Nereididae
parapodium	Paired, fleshy, lateral outgrowths from the body segments of polychaetes that (usually) bear chaetae
paratype	All type specimens in the type series which are not holotypes . Paratypes have no name-bearing function.
PC	Personal Computer / Principal Component in a Principal Component Analysis
PCA	Principal Component Analysis. A statistical method to describe multivariate data through its most prominent components
PCR	Polymerase Chain Reaction: a technology to amplify DNA
PDF	Portable Document Format: a proprietary file format to represent documents independently of the operating system or software used
pectinate chaetae	Comb-like chaetae
peristomium	The first body segment in polychaetes after the prostomium , surrounding the mouth
perl	A programming language
PERMANOVA	Permutational Multivariate Analysis of Variance: a statistical method, a non-parametric version of an analysis of variance using distance matrices and a Monte Carlo permutation technique to calculate probabilities
pharyngeal glands	A pair of glands next to the pharynx, occurring in some species in the polychaete family Syllidae
PHP	A programming (scripting) language used mainly for the dynamic creation of web sites
PID	Persistent Identifier: a permanent (long-lasting) reference to a digital object
pixel	The smallest unit of information in a digital image
Plazi	An initiative supporting and promoting openly accessible digital taxonomic literature, developing tools and repositories towards this end

	(http://plazi.org)
png	Portable Network Graphic: Image format
polychaetes	Bristle worms, belonging to the segmented worms of the class Annelida. Most polychaetes are marine and they are one of the most abundant groups of marine invertebrates
PRIMER	Plymouth Routines In Multivariate Ecological Research: a commercial statistical software package
projection image	Also called shadow image. Image that result from <u>CT</u> imaging, provide a greyscale side view of the specimen, with dense parts in darker shades and less dense parts in lighter shades of grey. From a series of projection images the cross-sections are calculated.
prostomium	Antermost region of the <u>polychaete</u> body, sometimes bearing eyes and / or antennae
proventricle	Heavily muscularised region of the digestive tract posterior to the pharynx, occurring in the <u>polychaete</u> family Syllidae
proximal	The point of an anatomical appendage which is nearest to the attachment to the main body
PTA	Phosphotungstic acid: an electron-dense stain used as a contrast agent in histology and <u>CT</u> imaging
pygidium	Posteriormost region of the <u>polychaete</u> body surrounding the anus
pyriform	Pear-shaped
R	A free and open-source statistical software package (http://www.r-project.org)
RAM	Random-access memory: a form of computer data storage for fast access
RDF	Resource Description Framework: A method to describe or model information in web resources, allowing computers to make "intelligent" use of this structured information. RDF is one of the main components of the <u>semantic web</u> .
reconstruction	Algorithmic process of transforming <u>projection images</u> into cross sections
REST API	A REST (Representational State Transfer) <u>API</u> uses <u>HTTP</u> methods and <u>URIs</u> as a means to communicate and execute operations between two machines.
ROI	Region of Interest: User-specified region of an image
RT-PCR	Real-Time Polymerase Chain Reaction: method to simultaneously amplify and quantify DNA
Scratchpads	A <u>virtual research environment</u> for the collaborative management and publication of biodiversity data. Based on the software <u>Drupal</u> (http://scratchpads.eu)
Sea Life Base	A global database of information on marine organisms, manually curated and therefore growing only slowly (http://www.sealifebase.org)
SEM	Scanning Electron Microscopy: a high-resolution imaging technique
semantic web	A term describing a web of data elements which are semantically

	annotated and can thus be exploited by computers
sequence data	Data comprising nucleic acid sequences
sigmoid	S-shaped
16S rRNA	16S ribosomal RNA, a gene often used in molecular species identification and the resulting phylogenetic studies
species complex	A group of <u>cryptic species</u>
species-id.net	A species information system based on the wiki software (http://species-id.net)
SQL	Structured Query Language, a programming language to manage and query data stored a relational database management system
stereo display	A way to display images conveying a three-dimensional impression
string	In computer programming: a sequence of symbols or numbers, often more specific: a data type storing text values
subacicular hooks	Stout <u>chaeta</u> in the form of a hook, exiting the <u>parapodial</u> lobe below the <u>acicula</u>
subdistal spine	Prominent denticle below the distal end of a <u>chaeta</u>
supra-acicular	Above the <u>acicula</u>
surface rendering, surface model	See <u>isosurface model</u>
TAU	Tel Aviv University Zoological Museum
taxonomic impediment	Term to express the current problems traditional <u>taxonomy</u> is facing: lack of readily available taxonomic information, gaps in taxonomic knowledge and shortage of experienced taxonomists
taxonomy	Scientific discipline that encompasses the description, naming, nomenclature, and classification of organisms
TaxPub	An <u>XML</u> extension for the <u>markup of taxonomic treatments</u> (http://sourceforge.net/projects/taxpub)
tentacular cirrus	Cirrus located on anterior cephalised segments of the <u>polychaete</u> body
text mining	Automatic extraction of targeted information from text
TGA	Truevision Graphics Adapter: an image format
3D	Three-dimensional
TraitBank	A recently launched initiative aiming at becoming a digital repository for information on species' life-cycles and interactions (http://www.eol.org/traitbank)
transfer values, transfer function	In <u>volume rendering</u> , a definition of the colour and opacity values which with each <u>voxel</u> should be displayed
TRY	A database on plant traits, most data only available upon request (http://www.try-db.org)
type material, type specimen	A biological reference specimen, usually kept in museums or biological collections, to which the scientific name of that specimen is formally attached
type species	The species to which the concept of its genus is permanently linked, used to define the genus

URI	Uniform Resource Identifier: a sequence of characters used to identify a web resource
URL	Uniform Resource Locator: often also synonymously used with “web address”, A URL is a <u>URI</u> which also provides a means to access the resource.
USNM	Smithsonian National Museum of Natural History
UTF-8	A text encoding scheme to represent characters in the Unicode character set (thus most of the world's writing systems)
VAXML	Virtual Anatomy eXtensible Markup Language: a <u>markup language</u> designed to facilitate the sharing of ‘virtual specimen’ data
ventral cirrus	<u>Cirrus</u> which (normally) arises from the lower part of the <u>neuropodium</u>
vertices	In geometry, a point describing the intersections of geometric shapes
ViBRANT	Virtual Biodiversity Research and Access Network for Taxonomy (http://www.vbrant.eu): an EU funded research project with a strong focus on cybertaxonomy and biodiversity informatics
virtual research environment	A web-based environment facilitating collaborative research
volume rendering	A way of displaying three-dimensional data acquired through <u>3D-imaging</u> techniques such as <u>CT</u> or <u>MRI</u> . Depending on the <u>transfer functions</u> applied, different densities of the sample can be visualised and /or colour-coded.
voxel	Volume pixel: the smallest unit of information in a three-dimensional digital space
VRE	See <u>virtual research environment</u>
VRML	A file format for representing <u>3D</u> objects
web service	A method for the communication between two machines or software programmes over the World Wide Web
Wikispecies	A project based on the wiki software aiming at creating a communal species information system (http://species.wikimedia.org)
WorMS	World Register of Marine Species (http://www.marinespecies.org): a database on marine species names, nomenclature and associated bibliographies.
XML	Extensible Markup Language: A <u>markup language</u> that defines a set of rules to structure data within documents, in a format which is readable by both humans and machines.
ZooBank	The official online registry of <u>taxonomic</u> and nomenclatural acts (and thus, species names) in zoology (http://www.zoobank.org)

Appendix I

Penev L, Blagoderov V, Georgiev T, Rycroft S, Scott B,
Faulwetter S (2011) ViBRANT deliverable 6.1: XML markup
tools and service. 10pp.



Deliverable 6.1: XML mark up tool and services

Leading partners: Pensoft, NHM

Compiled by:

Lyubomir Penev, Vladimir Blagoderov, Teodor Georgiev, Simon Rycroft, Benjamin Scott,
Sarah Faulwetter

November 2011

Explanation note

This document represents the workflow of creating an XML-tagged manuscript within the publishing module of Scratchpads, its export into the XML file, submission to ZooKeys, and its publication in three electronic versions: PDF, semantically enhanced HTML and XML.

The workflow is illustrated on the example of a new polychaete species description by Faulwetter et al. (2011), to be published in the special issue of ZooKeys commissioned by ViBRANT project at the end of November.

Description of the workflow:

1. An author creates a Publication project within a Scratchpad to which only a restricted set of users have access. The author(s) also provide additional information required by the article (e.g., title, author's details).
2. The author(s) prepare species pages (including descriptions, images, specimens etc.) within the Scratchpad. In case of a new taxon description author(s) use a temporary name (a placeholder). This placeholder acts as a surrogate for the final taxon name to ensure that the new name is not disclosed until the description has been accepted by the journal. The placeholder is linked (tagged) to data on their site, and the placeholder taxon name is linked to the final name. The author(s) select data to be included in the manuscript. Additional sections are added to the manuscript using a structure that will accommodate most taxonomic descriptions (Fig. 1) and images uploaded (Fig. 2.). Different stages of the manuscript preparation are illustrated on Fig. 3 When the preparation stage is complete, the author(s) preview the manuscript to make sure it is satisfactory (Fig. 4).
3. Author(s) submit the manuscript, which creates an archive of the manuscript components. The submission process automatically generates an XML representation of the document according to the TaxPub extension of the NLM/NCBI Journal Archiving DTD (<http://sourceforge.net/projects/taxpub/>). This document is automatically sent to the journal ZooKeys.
4. ZooKeys organises the peer review (see discussion on peer review). The reviewed paper, including reviewer's comments, is sent by e-mail back to the corresponding author.
5. Author(s) revise their manuscript and supporting data on their Scratchpad in response to the reviewers' comments.

Sphaerosyllis levantina sp. n. (Annelida: Polychaetes) from the eastern Mediterranean

View **Edit** Export Revisions Track XML

Metadata Taxa Default fields Sections Images Bibliography Privacy

Title	Operations
+ Introduction	Edit Remove
+ Material and methods	Edit Remove
+ Specimen collection and processing	Edit Remove
+ Morphometric analyses	Edit Remove
+ Electronic publication	Edit Remove
+ Results	Edit Remove
+ Taxonomic Results	Edit Remove
+ Sphaerosyllis levantina sp. nov.	Edit
+ Comparative material examined	Edit Remove
+ Type locality	Edit Remove
Sphaerosyllis sp. nov. taxon description Morphology	Edit term fields

Figure 1. Sections of the manuscript. Terms listed under taxon name correspond to the fields of the Species Profile Model (SPM) to be automatically included in the manuscript. Custom sections can be organised hierarchically.

Sphaerosyllis levantina sp. n. (Annelida: Polychaetes) from the eastern Mediterranean

View **Edit** Export Revisions Track XML

Metadata Taxa Default fields Sections Images Bibliography Privacy

Images

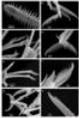
- + 
- + 
- + 
- + 

Figure 2. Selecting images to be included in the manuscript.

Sphaerosyllis levantina sp. n. (Annelida: Polychaetes) from the eastern Mediterranean

View Edit **Export** Revisions Track XML

Export publication:

- None -
- None - the publication export format.
Pensoft

Sphaerosyllis levantina sp. n. (Annelida: Polychaetes) from the eastern Mediterranean

View Edit **Export** Revisions Track XML

Export publication:

Pensoft
Please select the publication export format.

Your details

Title: *

First name: *

Surname: *

Initials: *

Institution

Affiliation:



polychaetes

Home » Groups » Syllidae Israel

Sphaerosyllis levantina sp. n. (Annelida: Polychaetes) from the eastern Mediterranean

View Edit Export Revisions Track XML

Edited by Sarah Faulwetter on Wed, 10/05/2011 - 07:32

Sphaerosyllis levantina sp. n. (Annelida: Polychaetes) from the eastern Mediterranean, with notes on character variation in *Sphaerosyllis hystrix* Claparède, 1963

Sarah Faulwetter¹, Georgios Chatzigeorgiou², Bella S. Galil³, Artemis Nicolaidou⁴, Christos Arvanitidis⁵

1. sanifa@hcmr.gr, Department of Zoology-Marine Biology, Faculty of Biology, National and Kapodestrian University of Athens, Panepistimiopolis, 15784, Athens, Greece
2. chatzigeorgiou@hcmr.gr, Department of Biology, University of Crete, 71409 Heraklion, Crete, Greece
3. bella@ocean.org.il, National Institute of Oceanography, Israel Oceanographic and Limnological Research, POB 8030, Haifa 31080, Israel
4. anikol@biol.aoa.gr, Department of Zoology-Marine Biology, Faculty of Biology, National and Kapodestrian University of Athens, Panepistimiopolis, 15784, Athens, Greece
5. arvanitidis@bioacademy.gr, Department of Zoology-Marine Biology, Faculty of Biology, National and Kapodestrian University of Athens, Panepistimiopolis, 15784, Athens, Greece

Figure 3. Different stages of the manuscript export process.

8. The manuscript and all supplementary data are unlocked on the Scratchpad and made public on the day of printed publication. At this time the placeholder taxon names are automatically substituted by the final published taxon name.

By default all Scratchpad data concerning the ZooKeys publication are kept private for steps 1 to 8 and made public at step 9, although the original taxon pages are normally public. However, the author(s) have the capacity to make all these data public from the outset.

Technical Implementation. A single Drupal module (called “Publication”) has been written to support the technical implementation of this workflow within the Scratchpads. This is available from the Scratchpad Subversion repository (<http://svn.scratchpads.eu/svn/scratchpads/trunk/modules/publication/>) along with other Scratchpad project written dependencies. Software dependencies include the Drupal community’s Organic Groups module (<http://drupal.org/project/og>) and Content Construction Kit (<http://drupal.org/project/ckk>) modules, in addition to the Scratchpad project’s Species Profile Module (SPM) and Taxonomy Tree modules. The Publication module provides a new Drupal content-type (also called “Publication”) that is set to be an “Organic Group”. This enables an author to assign other users to a publication object and optionally restrict access to content associated with that publication. The Publication module creates three other simple content types that are used to provide additional sections for the publication. The first of these supports general sections common to most publications (e.g. Discussion, Materials and Methods) and taxon specific sections that allow users to add sections to each taxon treatment (e.g. Citations, Type Material). The second of these enables users to control which data fields appear in each taxon treatment and their relative order in the text. Finally, an image caption content type is provided to enable users to annotate their images.

In summary the Publication module provides an intuitive interface that allows users to select and order content from their site and associate this with the publication, providing a many-to-many link between publication objects and other content types (e.g. Image, Bibliography). Thus for example, a single image can be used in many publications, and a single publication can have many images. The module also supports the communication between the user’s Scratchpad and the publisher transferring the TaxPub XML representation of the manuscript to ZooKeys during submission, revision and final acceptance. TaxPub is an extension of the National Library of Medicine (NLM) / National Center for Biotechnology Information (NCBI) Journal Archiving Document Type Definition (DTD) for the markup of taxonomic treatments.

Sphaerosyllis levantina sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in *Sphaerosyllis hystrix* Claparède, 1863

Sarah Faulwetter^{1,4†}, Georgios Chatzigeorgiou^{2,4‡}, Bella S. Galil^{3§},
Artemis Nicolaidou^{1,1}, Christos Arvanitidis^{4¶}

1 Department of Zoology-Marine Biology, Faculty of Biology, National and Kapodestrian University of Athens, Panepistimiopolis, 15784, Athens, Greece **2** Department of Biology, University of Crete, 71409 Heraklion, Crete, Greece **3** National Institute of Oceanography, Israel Oceanographic & Limnological Research, POB 8030, Haifa 31080, Israel **4** Institute of Marine Biology and Genetics, Hellenic Centre for Marine Research, 71003 Heraklion, Crete, Greece

Sphaerosyllis levantina sp. n. (Annelida) from the eastern Mediterranean...

5

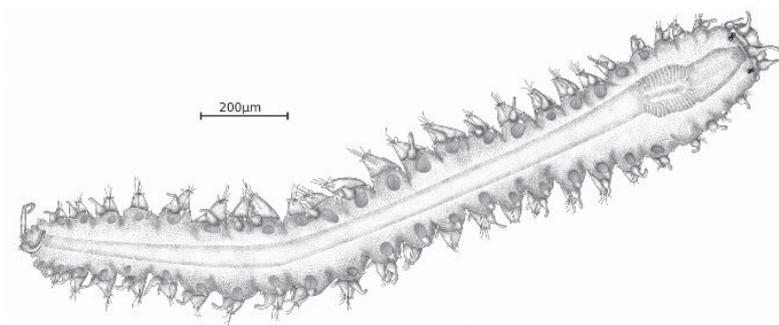


Figure 1. *Sphaerosyllis levantina* sp. n. holotype, dorsal view

lenic Centre for Marine Research, Anavyssos, Greece; Chalkida, Aegean Sea, Greece: 1 specimen [Label: 56 – *Sphaerosyllis hystrix*, κατώτερη μεσοπαλαική Χαλκίδας, Στενά Ευρίπου, Ξενοδοχείο Λούσι, St. 18, 25.9.97 0-0.5m, Άτομα: 1, Διδακτορικού Μίλτου] (= lower intertidal zone, Chalkida, Eviros Strait, Hotel Lousi, coll. M.S. Kitsos), Chalkida, Aegean Sea, Greece: 1 specimen [Label: 26 – *Sphaerosyllis hystrix*, κατώτερη

Taxon treatments	Taxon names
------------------	-------------

[Top](#) | [Abstract](#) | [Introduction](#) | [Material and methods](#) | [Results](#) | [Discussion](#) | [Supplementary material](#) | [Acknowledgements](#) | [References](#)

ZooKeys @@@: @-@, doi: 10.3897/zookeys.@@.1877

***Sphaerosyllis levantina* sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in**

Sarah Faulwetter^{1,4,†}, Georgios Chatzigeorgiou^{2,4,†}, Bella S. Galil^{3,†}, Artemis Nicolaidou^{1,†}, Christos Arvanitidis^{4,†}

1 Department of Zoology-Marine Biology, Faculty of Biology, National and Kapodestrian University of Athens, Panepistimiopolis, 15784, Athens, Greece
 2 Department of Biology, University of Crete, 71409 Heraklion, Crete, Greece
 3 National Institute of Oceanography, Israel Oceanographic & Limnological Research, POB 8030, Haifa 31080, Israel
 4 Institute of Marine Biology and Genetics, Hellenic Centre for Marine Research, 71003 Heraklion, Crete, Greece

† [urn:lsid:zoobank.org:author:](https://doi.org/10.3897/zookeys.@@.1877)
 † [urn:lsid:zoobank.org:author:](https://doi.org/10.3897/zookeys.@@.1877)
 † [urn:lsid:zoobank.org:author:](https://doi.org/10.3897/zookeys.@@.1877)
 † [urn:lsid:zoobank.org:author:](https://doi.org/10.3897/zookeys.@@.1877)
 † [urn:lsid:zoobank.org:author:](https://doi.org/10.3897/zookeys.@@.1877)

Corresponding author: Sarah Faulwetter (sarifa@hcmr.gr)

Academic editor: C. Glasby

Received 3 August 2011 | Accepted 18 October 2011

(C) Sarah Faulwetter. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits medium, provided the original author and source are credited.

For reference, use of the paginated PDF or printed version of this article is recommended.

Abstract

Examination of polychaete specimens from Haifa Bay (Israel, eastern Mediterranean Sea) revealed several individuals exhibiting morphological characters similar to *Sphaerosyllis* Claparède, 1863. A detailed morphometrical analysis of the Israeli specimens in comparison to specimens of *S. hystrix* and *S. boeroi* Musco of the former as a new species, *S. levantina* sp. n. Individuals of *S. hystrix* formed a very heterogeneous group with strong character variation. The distribution of the species is discussed based on literature records.

Keywords

Polychaetes, Syllidae, Exogoninae, *Sphaerosyllis*, new species, Mediterranean, Cybertaxonomy, Scratchpads

Introduction

The polychaete genus *Sphaerosyllis* Claparède, 1863 (Annelida) is one of the most species-rich genera of the syllid subfamily Exogoninae. The genus *Sphaerosyllis* after the recent split of the group into the three genera *Sphaerosyllis*, *Prospaerosyllis* and *Erinaceosyllis* (San Martín 2005). *S. levantina* sp. n. is described from the Mediterranean Sea (Musco and Gianrande 2005). one of them described but yet unnamed (San Martín 2003). another one in the

Figure 6. Final proofs of the manuscript after peer-review (HTML)

```

1 <?xml version="1.0"?>
2 <article xmlns:xlink="http://www.w3.org/1999/XMLSchema" xmlns="http://www.w3.org/1999/Math/MathML" xmlns:xlink="http://www.w3.org/1999/xlink"
3   xmlns:xs="http://www.w3.org/2001/XMLSchema-instance" xmlns:tp="http://www.plazi.org/taxpub">
4   <front>
5     <journal-meta>
6       <journal-id journal-id-type="publisher-id">ZooKeys</journal-id>
7       <journal-title-group>
8         <journal-title>ZooKeys</journal-title>
9         <abbrev-journal-title>ZooKeys</abbrev-journal-title>
10      </journal-title-group>
11      <issn pub-type="ppub">1318-2969</issn>
12      <issn pub-type="pub">1318-2970</issn>
13      <publisher>
14        <publisher-name>Pensoft Publishers</publisher-name>
15      </publisher>
16    </journal-meta>
17    <article-meta>
18      <article-id pub-id-type="doi">10.3897/zookeys.89.1877</article-id>
19      <title-group>
20        <article-title>
21          <title>Sphaeronyllia levantina (Italo) sp. n. (Annelida) from the eastern Mediterranean, with notes on character variation in
22          <title>Sphaeronyllia hystrix (Italo) Claparède, 1863
23        </title>
24      </title-group>
25      <contrib-group>
26        <contrib contrib-type="author" xlink:type="simple">
27          <name name-style="western">
28            <surname>Lettier</surname>
29            <given-names>Sarah</given-names>
30          </name>
31          <xref ref-type="aff" rid="A1"></xref>
32          <xref ref-type="aff" rid="A2"></xref>
33          <xref content-type="text" xlink:type="simple">urn:lsid:zoobank.org:author:</xref>
34        </contrib>
35        <contrib contrib-type="author" xlink:type="simple">
36          <name name-style="western">
37            <surname>Chatzigeorgiou</surname>
38            <given-names>Georgios</given-names>
39          </name>
40          <xref ref-type="aff" rid="A2"></xref>
41          <xref ref-type="aff" rid="A1"></xref>
42          <xref content-type="text" xlink:type="simple">urn:lsid:zoobank.org:author:</xref>
43        </contrib>
44        <contrib contrib-type="author" xlink:type="simple">
45          <name name-style="western">
46            <surname>Belló</surname>
47            <given-names>Belia S.</given-names>
48          </name>

```

Figure 7. Final proofs of the manuscript after peer-review (XML)

Appendix II

Documentation of the data entry interface of the
polytraits database

Overview of the *polytraits* data entry interface



The screenshot shows the login page for the polytraits data entry interface. At the top left is the polytraits logo, and at the top right is the title "data entry interface - login". Below the title is a navigation bar with links for Home, Taxa, Traits, Download data (with a dropdown arrow), and Login. The main content area contains a login form with two input fields: "user:" and "pass:", each followed by a text box. Below these fields is a "login" button. A disclaimer is located at the bottom right of the page, stating: "Disclaimer: The data entry interface works well with Firefox and other modern browsers. There might be some glitches when used with Internet Explorer."

Figure 1: The data entry interface is password protected to ensure only authorised persons can enter information.



The screenshot shows the traits overview page. At the top is a teal header with the text "trait database | list of traits & their modalities". Below the header is a navigation bar with links for Traits, Taxa, References, Statistics, Tools, and Help. The main content area is titled "Adult traits" and lists several traits, each with a "+" sign to expand it. The traits listed are: Body size (max), Complex species, Depth zonation (benthos), Depth zonation (pelagic), Ecosystem engineering, Feeding structure, Feeding type, Habitat type, Intra- and interspecific competition, and Lifespan. Each trait header has a link "taxa for this trait" with a small icon. The "Complex species" trait is expanded, showing two text boxes: "yes" with the text "Complex species reported in the literature." and "no" with the text "No complex species reported in the literature." Below these text boxes is a link "add new value here".

Figure 2: Traits overview page. Each trait can be expanded by clicking on the “+” sign, providing information on the modalities of the traits. Traits, modalities and their definitions can be added and modified through this interface. On the right side of each trait header, a link “Taxa for this trait” lists all taxa and the information which has been collected for this trait.

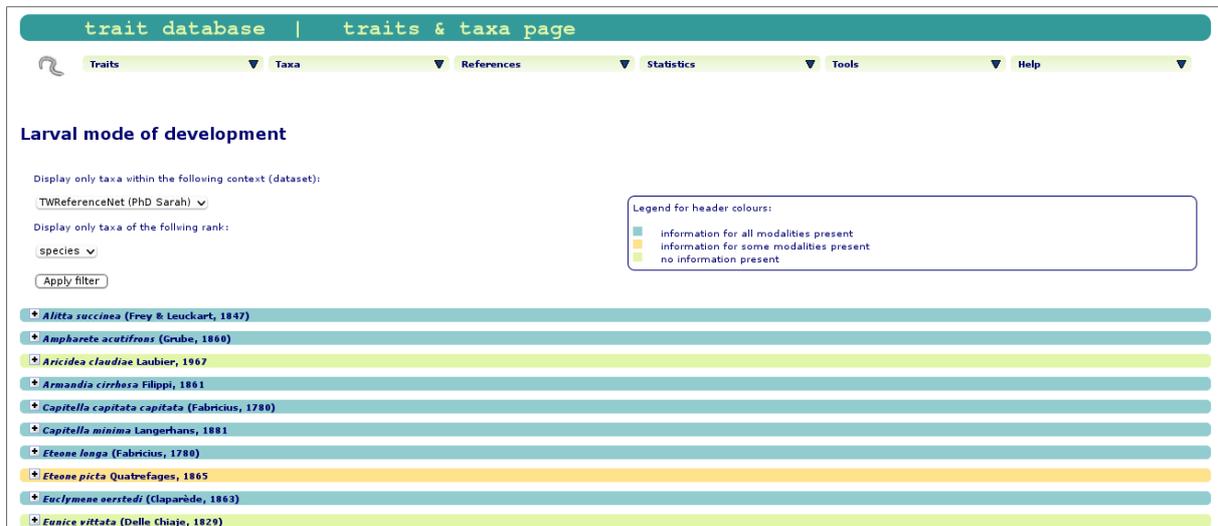


Figure 3: Overview page of taxa per trait. Each taxon heading can be expanded by clicking on the “+” sign, revealing the modalities for this trait and the information gathered for the taxon. Colour-coded headers provide information on the completeness of information for each taxon: blue=information for all modalities present, yellow=information for some modalities present, green=no information present. The displayed taxa can also be filtered according to rank and pre-defined datasets they occur in (see also Figure 11).



Figure 4: Taxa present in the system can be browsed in various ways. The figure depicts alphabetical sorting of all taxa, divided into the ranks of species, genus and family.

trait database | all currently valid taxa, alphabetically

Objective (homotypic) synonyms and misspellings are displayed in grey. Clicking on them will lead to the currently accepted taxon. Subjective (heterotypic) synonyms are displayed normally and have their own taxon page but their currently accepted name is displayed next to them

Species level taxa	Genus level taxa	Family level taxa
Acrocirrus frontifilis (Grube, 1860)	Acrocirrus Grube, 1873	Aberrantidae Wolf, 1987
Adercodon pleijeli Mackie, 1993	Adercodon Mackie, 1993	Acoetidae Kinberg, 1856
Aglaphamus agilis (Lanqerhans, 1880)	Aglaphamus Kinberg, 1865	Accoriridae Bancs, 1969
<i>Aglaphamus rubella</i> (Michaelsen, 1897) (subjective synonym according to http://www.marinespecies.org)	Alitta Bakken & Wilson, 2005	Aeolosomatidae
Aglaphamus circinata (Verrill in Smith & Harver, 1874)	Alkmaria Horst, 1919	Alciopidae Ehlers, 1864
Alitta succinea (Frey & Leuckart, 1847)	Amaeana Hartman, 1959	Alvinellidae Desbruyères & Laubier, 1986
<i>Nereis succinea</i> Frey & Leuckart, 1847 (objective synonym)	Amaea	Ampharetidae
<i>Nereis (Neanthes) succinea</i> (Frey & Leuckart, 1847) (objective synonym)	Amblonyxillus Grube, 1857	Amphinomidae Savigny in Lamarck, 1818
<i>Neanthes succinea</i> (Frey & Leuckart, 1847) (objective synonym)	Ampharete Malmgren, 1866	Anaplosyllinae
<i>Nereis reibischi</i> Heinen, 1911 (subjective synonym according to Hartmann-Schröder, G. (1996))	Amphicorina Quatrefages, 1850	Antonbruunidae
<i>Nereis belavanensis</i> Pflugfelder, 1933 (subjective synonym according to http://www.marinespecies.org)	Amphicteis	Aphroditidae Malmgren, 1867
<i>Nereis (Neanthes) saltoni</i> Hartman, 1936 (subjective synonym according to http://www.marinespecies.org)	Amphicteus Savigny, 1818	Apistobranchidae Mesnil & Caullery, 1898
<i>Nereis (Alitta) oxyroda</i> Marenzeller, 1879 (subjective synonym according to Wilson, R.S. (1988))	Amphiglena	Aroncolidae
<i>Neanthes perrieri</i> Saint-Joseph, 1898 (subjective synonym)		
<i>Nereis glandulosa</i> Ehlers, 1908 (subjective synonym according to http://www.marinespecies.org)		

Figure 5: Taxa present in the system can be browsed in various ways. The figure depicts taxa grouped by their currently accepted name, listing synonyms, why they are not currently accepted (e.g. homotypic synonym) and the source for the synonymy.

trait database | species classification

- Polychaeta
 - Polychaeta Palpata
 - Eunicida
 - Dorvilleidae
 - Schistomeringos
 - Schistomeringos rudolphi* (Chiaje, 1828) [lexical variant]
 - Schistomeringos rudolphi* (Delle Chiaje, 1828)
 - Schistomeringos neglecta* (Fauvel, 1923)
 - Schistomeringos caeca* (Webster & Benedict, 1887)
 - Dorvillea rudolphi* (Delle Chiaje, 1828) [objective synonym]
 - Protodorvillea Pettibone, 1961
 - Dorvillea Parfitt, 1856
 - Ophryotrocha Claparède & Meczniow, 1869
 - Ikosipodus Westheide, 1982
 - Pettiboneia Orensanz, 1973
 - Eunicidae
 - Lumbrineridae
 - Oeonidae
 - Hartmaniellidae
 - Histiobdellidae
 - Ichthyotomidae
 - Onuphidae
 - Phyllodocida
 - Polysarchoidea

Figure 6: Taxa present in the system can be browsed in various ways. The figure depicts the biological classification of taxa. A clade can be expanded by clicking on the “+” sign, displaying its descendants.

trait database | Alitta succinea - taxon page

Navigation: Traits | Taxa | References | Statistics | Tools | Help

Alitta succinea (Frey & Leuckart, 1847)

Includes:
Heanthes succinea (Frey & Leuckart, 1847) (according to <http://www.marinespecies.org>)
Nereis succinea Frey & Leuckart, 1847 (according to [Hartmann-Schröder, G. \(1996\)](http://www.marinespecies.org))
Nereis (Heanthes) succinea (Frey & Leuckart, 1847) (according to <http://www.marinespecies.org>)

Currently synonymized but treated on separate taxon pages:
Nereis relicchi Heinen, 1911 (according to [Hartmann-Schröder, G. \(1996\)](http://www.marinespecies.org))
Heanthes perrieri Saint-Joseph, 1898
Nereis (Heanthes) australis Treadwell, 1923 (according to <http://www.marinespecies.org>)
Nereis (Heanthes) saltoni Hartman, 1936 (according to <http://www.marinespecies.org>)
Nereis acutifolia Ehlers, 1901 (according to <http://www.marinespecies.org>)
Nereis indusianensis Pflugfelder, 1933 (according to <http://www.marinespecies.org>)
Nereis glandulosa Ehlers, 1908 (according to <http://www.marinespecies.org>)
Nereis limbata Ehlers, 1868 (according to <http://www.marinespecies.org>)
Nereis (Alitta) oxyptoda Narenzeller, 1879 (according to [Wilson, R.S. \(1988\)](http://www.marinespecies.org))
Nereis alatopalis Wesenberg-Lund, 1949 (according to [Wilson, R.S. \(1988\)](http://www.marinespecies.org))

Legend for header colours:
 information for all modalities present
 information for some modalities present
 no information present

Completeness of info inferred from related taxa:
 SYN = info from synonyms
 GEN = info from any member of the genus
 FAM = info from any member of the Family

Source info:
 Raw info present
 No raw info present

Adult traits

Trait	Completeness (inferred info):	SYN	GEN	FAM
Body size (max)	Information for some modalities present			
Complex species	Information for all modalities present			
Depth zonation (benthos)	Information for all modalities present			
Depth zonation (pelagic)	Information for all modalities present			
Ecosystem engineering	Information for all modalities present			

Figure 7: A taxon page, listing all traits for the taxon. The page also displays synonyms of the current taxon. Headers for each trait can be expanded by clicking on the “+” sign, displaying the modalities of the trait. Colour-coding of headers as in Figure 3.

Complex species: Completeness (inferred info): SYN GEN FAM

Depth zonation (benthos): Completeness (inferred info): SYN GEN FAM

Modality	present	absent	unknown
supralittoral zone	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
littoral zone	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Information inferred from other taxa in this family:
[Hadiste diversicolor](#) : present
[Perineris cultifera](#) : present
[Nereididae](#) : present

Fong, P.P. (1991) The effects of salinity, temperature, and photoperiod on epitokal metamorphosis in *Heanthes succinea* (Frey et Leuckart) from San Francisco Bay. *Journal of Experimental Marine Biology and Ecology*, 149:177-190. [added by Christina Pavloudi on 2012-04-10 19:41:08]

Hartmann-Schröder, G. (1996) *Annelida, Borstenwürmer, Polychaeta*. Gustav Fischer Verlag, Jena. 648pp. [added by Sarah Faulwetter on 2009-10-19 15:59:44]

Imajima, M.(1991) *Catalogue of polychaetous annelids: Information retrieval on species of the family Hareidae*. National Science Museum, Tokyo. [added by Christina Pavloudi on 2012-04-10 23:39:37]

Vieitez, J., Alos, C., Parapan, J., Besteiro, C., Moreira, J., Núñez, J., Laborda, A., San Martín, G. (2004) *Fauna Iberica, Vol. 25. Annelida Polychaeta I*. Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid. 1-530pp. [added by Sarah Faulwetter on 2012-09-28 13:06:50]

Ergen, Z., Çınar, M.E., Daglı, E., Güley, K. (2006) Seasonal dynamics of soft-bottom polychaetes in Izmir Bay (Aegean

Figure 8: An expanded trait header, showing the modalities and information present for each modality. Each modality can be assigned three values: present, absent or unknown, the latter being default. When a value (present or absent) is assigned to a value, a literature source has to be specified (see Figure 9). Literature entries can be modified by clicking on the pencil icon. The information on when the information was added and by whom are displayed next to each literature entry. For modalities with no information present, the system displays existing information on other taxa in the same family on the right side of the page. This can help to detect outliers in the data and thus helps during quality control. For an explanation of icons see Figures 10, 16, 17 and 18.

Reproductive traits

- Age at first reproduction
- Developmental mechanism
- Egg size
- Epitoky
- Factors triggering reproduction
- Fecundity

1 - 50	<input type="checkbox"/> present
	<input type="checkbox"/> absent
	<input checked="" type="checkbox"/> unknown
50 - 500	<input checked="" type="checkbox"/> present
	<input type="checkbox"/> absent
	<input checked="" type="checkbox"/> unknown
500 - 2500	<input type="checkbox"/> present
	<input type="checkbox"/> absent
	<input checked="" type="checkbox"/> unknown
	<input type="checkbox"/> present

pick a reference

author contains: Pleijel search!

or enter a new one.

- Mackie, A.S.Y., Pleijel, F., Rouse, G.W. (2005) Revision of Aberranta Hartman, 1965 (Aberrantidae: Annelida), with descriptions of new species from the Mediterranean and Hong Kong. *Marine Ecology*, 26:197-208.
- Pleijel, F. (1991) Phylogeny and classification of the Phyllodocidae (Polychaeta). *Zoologica Scripta*, 20:225-261.
- Pleijel, F. (1993) Polychaeta Phyllodocidae. *Marine Invertebrates of Scandinavia*, 8:1-158. **0456**
- Pleijel, F., Mackie, A.S.Y. (1993) *Nereiphylla pusilla* (Polychaeta, Phyllodocidae) rediscovered and redescribed from Sicily. *Vie et Milieu*, 43:155-160.
- Rouse, G.W., Pleijel, F. (2001) *Polychaetes*. Oxford University Press, Oxford. 354pp.
- Rouse, G.W., Pleijel, F. (2006) *Reproductive Biology and Phylogeny of Annelida*. Science Publishers, New Hampshire. 688pp.
- Rousset, V., Rouse, G.W., Siddall, M.E., Tillier, A., Pleijel, F. (2004) The

ok cancel

Figure 9: Entering a value for a modality. When ticking any of the boxes for “present” or “absent”, a dialogue window opens which allows to search for a literature reference stored in the system. If no reference is found according to the search criteria, a link leads to a separate literature entry interface. A reference is selected by ticking the box in front of it. By clicking “OK” it is then assigned to the value for the respective modality and taxon.

Adult traits

- Body size (max)
- Complex species
- Depth zonation (benthos)

supralittoral zone	<input type="checkbox"/> present
	<input type="checkbox"/> absent
	<input checked="" type="checkbox"/> unknown
littoral zone	<input checked="" type="checkbox"/> present
	<input type="checkbox"/> absent
	<input type="checkbox"/> unknown

Nereis (Neanthes) saltoni Hartman, 1936 (according to <http://www.marinespecies.org>)
Nereis acutifolia Ehlers, 1901 (according to <http://www.marinespecies.org>)
Nereis belawanensis Huggfelder, 1923 (according to <http://www.marinespecies.org>)
Nereis glandulosa Ehlers, 1900 (according to <http://www.marinespecies.org>)
Nereis limbata Ehlers, 1868 (according to <http://www.marinespecies.org>)
Nereis (Alitta) oxygoda Harenzeller, 1879 (according to <http://www.marinespecies.org>)
Nereis slatopalis Wesenberg-Lund, 1949 (according to <http://www.marinespecies.org>)

Fong, P.P. (1991) The effects of salinity, temperature, and photoperiod on epitokal metamorphosis in *Neanthes succinea* (Frey et Leuckart) from San Francisco Bay. *Journal of Experimental Marine Biology and Ecology*, 149:177-190. [added by Christina Pavloudi on 2012-04-10 19:41:08]

Hartmann-Schröder, G. (1996) *Annelida, Borstenwürmer, Polychaeta*. Gustav Fischer Verlag, Jena. 648pp. [added by Sarah Faulvetter on 2009-10-19 15:59:44]

Imajima, M. (1991) Catalogue of polychaetous annelids: Information retrieval on species of the family Hereidae. National Science Museum, Tokyo. [added by Christina Pavloudi on 2012-04-10 23:39:37]

Vieitez, J., Alos, C., Parapar, J., Besteiro, C., Moreira, J., Núñez, J., Laborda, A., San Martín, G. (2004) *Fauna Iberica*. Vol. 25. *Annelida Polychaeta I*. Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid. 1-530pp. [added by Sarah Faulvetter on 2012-09-28 13:06:50]

polytraits.lifewatchgreece.eu/data entry interface

Hartmann-Schröder, G. (1996) *Annelida, Borstenwürmer, Polychaeta*. Gustav Fischer Verlag, Jena. 648pp.

p.208: "Die vertikale Verbreitung reicht vom oberen Eulitoral bis ins obere, seltener bis ins untere Sublitoral."

added / modified by Sarah Faulvetter on 2012-06-12 15:32:43

Edit/Add

Nereis: present
Medusa diversicolor: present
Paranereis cultrifera: present
Hereidae: present

Figure 10: The original text passage which lead to the assignment of the information is recorded by clicking on the little book icon next to each reference. A separate dialogue window opens providing a free text field for the text passage. When no original data is present for a literature entry for a specific modality, the book icon is crossed out.

trait database | enter new taxon

Traits Taxa References Statistics Tools Help

Enter new taxon

The classification in this database is rank-based. This means that every species has to be associated with a rank. Most ranks, including sub- super - and infra-categories can be stored by the database, however, the automatic inference of data will only work for the six classical ranks species - genus - family - order - class and phylum. You are therefore encouraged to enter only taxa of one of these categories in order to prevent data loss and inconsistencies.

enter the new taxon you would like to enter into the database (just taxon name, no authority)

Taxon:

enter

Enter synonym

You can enter synonyms for taxa existing already in the database. Please provide the source of the synonymy.

Synonym to be entered:

Taxon: Authority:

Valid name for this synonym:

- Vermilopis monodiscus Zibrowius, 1968
- Vermilopis striaticeps (Grube, 1862)
- Virchowia Langerhans, 1879
- Virchowia clavata Langerhans, 1879

Figure 11: Taxon entry interface. New taxa and synonyms can be entered into the system by providing the taxonomic name. The system will automatically receive the full information for the taxon from the World Register of Marine Species and insert the taxon in the right position in the classification. Synonyms or alternative names for taxa already existing in the database can likewise be entered. The source of synonymy needs to be specified.

trait database | contexts

Traits Taxa References Statistics Tools Help

Add a new context

Remove an existing context

Add taxa to an existing context

Remove taxa from an existing context

Choose the context to which you want to add taxa:

TWReferenceNet (PhD Sarah) ▾

Enter all taxa (without authority) that belong to this new context (one taxon per line):

Eupicidae

Figure 12: Context management interface. Taxa can be grouped to “contexts”, if e.g. information for a specific set of taxa for a specific analysis need to be gathered. This allows for displaying only a subset of taxa on the taxon overview pages, giving a better overview.

trait database publications	
Traits	Taxa
References	Statistics
Tools	Help
search for publications: author contains: <input type="text"/> search! (leave empty to show all)	
Abbiati, M., Bianchi, C.N., Castelli, A. (1987) Polychaete vertical zonation along a littoral cliff in the Western Mediterranean. <i>Marine Ecology</i> , 8:33-48.	full reference trait data
Abbiati, M., Maltagliati, F. (1992) Genetic population structure of <i>Neanthes succinea</i> (Polychaeta: Nereididae). <i>Journal of the Marine Biological Association of the United Kingdom</i> , 72:511-517.	full reference trait data
Abd-Elnaby, F.A. (2009) New Records of Polychaetes from the South Part of Suez Canal, Egypt. <i>World Journal of Fish and Marine Sciences</i> , 1:7-19.	full reference trait data
Abd-Elnaby, F.A. (2009) Polychaete Study in Northeastern Mediterranean Coast of Egypt. <i>World Journal of Fish and Marine Sciences</i> , 1:85-93.	full reference trait data
Aceró, I., San Martín, G. (1986) Poliquetos epibiontes del primer horizonte de algas fotofilas en las provincias de Cadiz y Malaga. Estudio faunístico comparado. <i>Boletín de la Real Sociedad Española de Historia Natural</i> , 81:5-24.	full reference trait data
Aguado, M.T., Nygren, A., Siddall, M.E. (2007) Phylogeny of Syllidae (Polychaeta) based on combined molecular analysis of nuclear and mitochondrial genes. <i>Cladistics</i> , 23:552-564.	full reference trait data

Figure 13: Literature management system. Entries can be browsed, searched and modified. A link “trait data” on the right side lists all information present in the database for that entry (Figure 15).

trait database Coverage of original data				
Traits	Taxa			
References	Statistics			
Tools	Help			
reference	nr of entries (total: 25108) ▼	with original data (total: 20685)	without original data (total: 4423)	% completed (total: 82.4)
Rouse, G.W., Pleijel, F. (2001) Polychaetes. Oxford University Press, Oxford, 354pp.	5466	5466	0	100
Rouse, G.W., Pleijel, F. (2006) Reproductive Biology and Phylogeny of Annelida. Science Publishers, New Hampshire, 688pp.	2252	2252	0	100
Hartmann-Schröder, G. (1996) Annelida, Borstentwürmer, Polychaeta. Gustav Fischer Verlag, Jena, 648pp.	1755	1754	1	99.9
Almeida, W.D.O., Christoffersen, M.L., Amorim, D.D.S., Garaffoni, A.R.S., Silva, G.S. (2003) Polychaeta, Annelida and Articulata are not monophyletic: articulating the Metameria (Metazoa, Coelomata). Revista Brasileira de Zoologia, 20:23-57.	1540	1540	0	100
Expert's Judgement	1174	803	371	68.4
BIOTIC - Biological Traits Information Catalogue. http://www.marlin.ac.uk/biotic/	530	324	206	61.1
Eckelbarger, K.J. (2005) Oogenesis and Oocytes. Hydrobiologia, 535/536:179-198.	502	497	5	99
Borja, A., Franco, J., Pérez, V. (2000) A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos within European Estuarine and Coastal Environments. Marine Pollution Bulletin, 40:1100-1114.	491	1	490	0.2
Fauchald, K., Jumars, P.A. (1979) The Diet of Worms: A Study of Polychaete Feeding Guilds. Oceanography and Marine Biology: Annual Review, 17:193-284.	452	327	125	72.3
Arvanitidis, C. (1994) Systematic and bionomic study of the macrobenthic Polychaetes (Annelida) on the northern Aegean. PhD Thesis, Aristotelian University of Thessaloniki, 512 pp.	450	220	230	48.9
	410	388	22	94.6

Figure 14: Summary view of all literature sources and the number of entries with and with original text passage for each reference. By clicking on the title of the reference, a list of all entries for this reference is displayed (Figure 15).

trait database | Entries per reference

Traits Taxa References Statistics Tools Help

Rouse, G.W., Pleijel, F. (2001) Polychaetes. Oxford University Press, Oxford. 354pp.

Display only taxa within the following context (dataset):

5466 entries found. Click on table headers to sort them.

taxon	rank	trait	modality	value	original data	added by
Aberrantidae	family	Depth zonation (benthos)	bathyal zone	present	   	Sarah Faulwetter on 2010-09-22 12:17:05
Aberrantidae	family	Depth zonation (benthos)	sublittoral zone	present	   	Sarah Faulwetter on 2010-09-22 12:15:25
Acoetidae	family	Depth zonation (benthos)	abyssal zone	absent	   	Sarah Faulwetter on 2010-09-24 10:33:22
Acoetidae	family	Depth zonation (benthos)	bathyal zone	present	   	Sarah Faulwetter on 2010-09-22 12:19:48
Acoetidae	family	Depth zonation (benthos)	hadal zone	absent	   	Sarah Faulwetter on 2010-09-24 10:33:33
Acoetidae	family	Depth zonation (benthos)	littoral zone	present	   	Sarah Faulwetter on 2010-09-22 12:19:31
Acoetidae	family	Depth zonation (benthos)	sublittoral zone	present	   	Sarah Faulwetter on 2010-09-22 12:19:38
Acoetidae	family	Depth zonation (benthos)	supralittoral zone	absent	   	Sarah Faulwetter on 2010-09-24 10:33:51
Ameritidae	family	Depth zonation (benthos)	abyssal zone	present	   	Sarah Faulwetter on 2010-09-22

Figure 15: Information per reference. The table columns can be sorted by clicking on the header. Four different icons allow for additional actions: the book icon opens a dialogue to add or edit the original text passage (see Figure 10), the tree icon opens a dialogue allowing the user to assign the data to another taxon (Figure 16), the two arrows allow the user to assign the data to an different modality (Figure 17) and the blue arrow opens a dialogue allowing the user to automatically assign the information to all or selected child taxa (Figure 18).

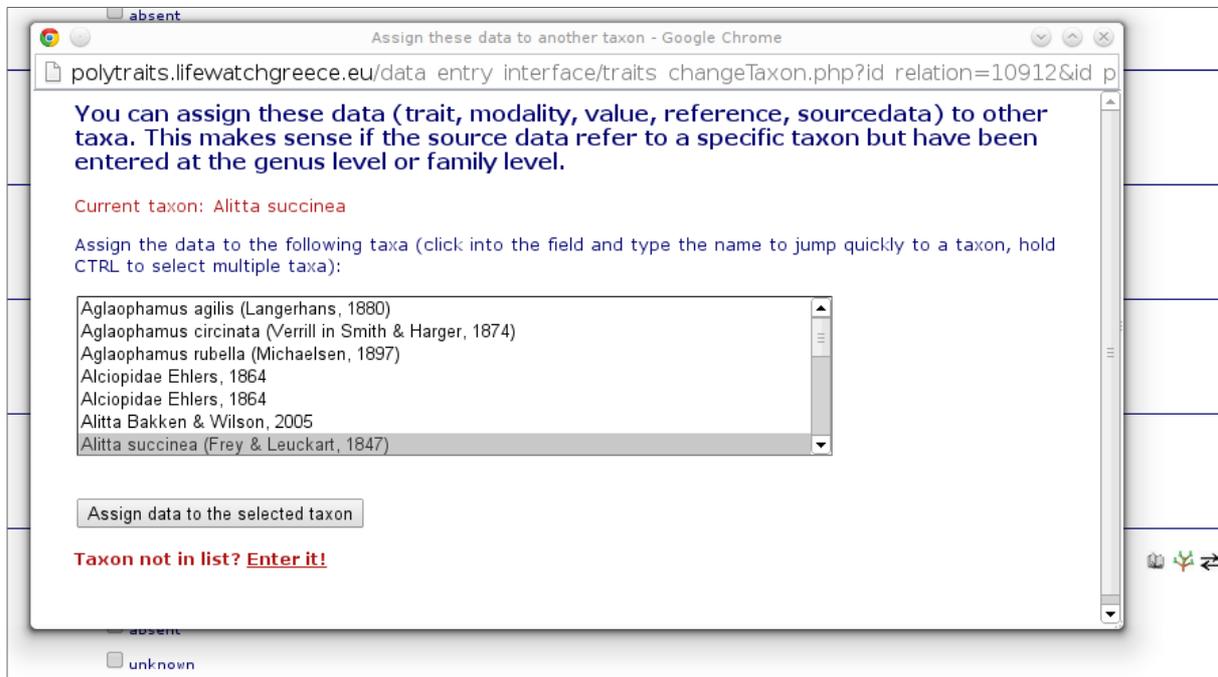


Figure 16: By clicking on the tree icon, a dialogue opens allowing the user to assign the information to a different taxon.

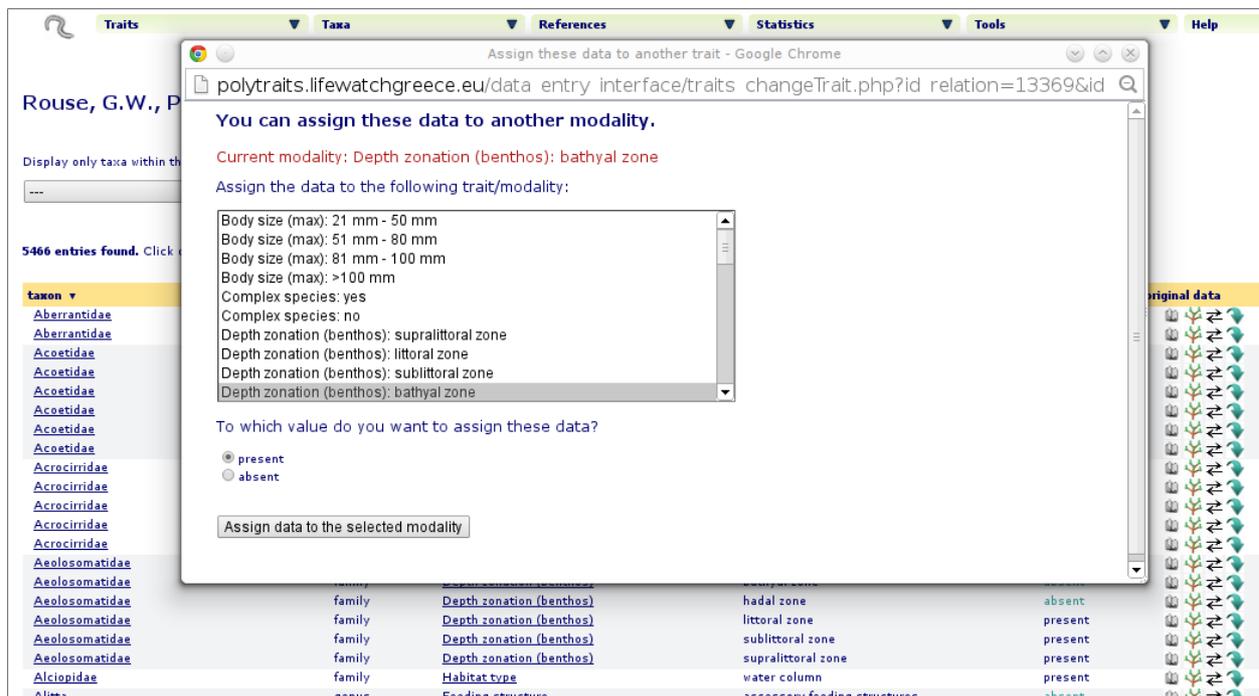


Figure 17: By clicking on the two black arrows, a dialogue opens allowing the user to assign the information to a different taxon.

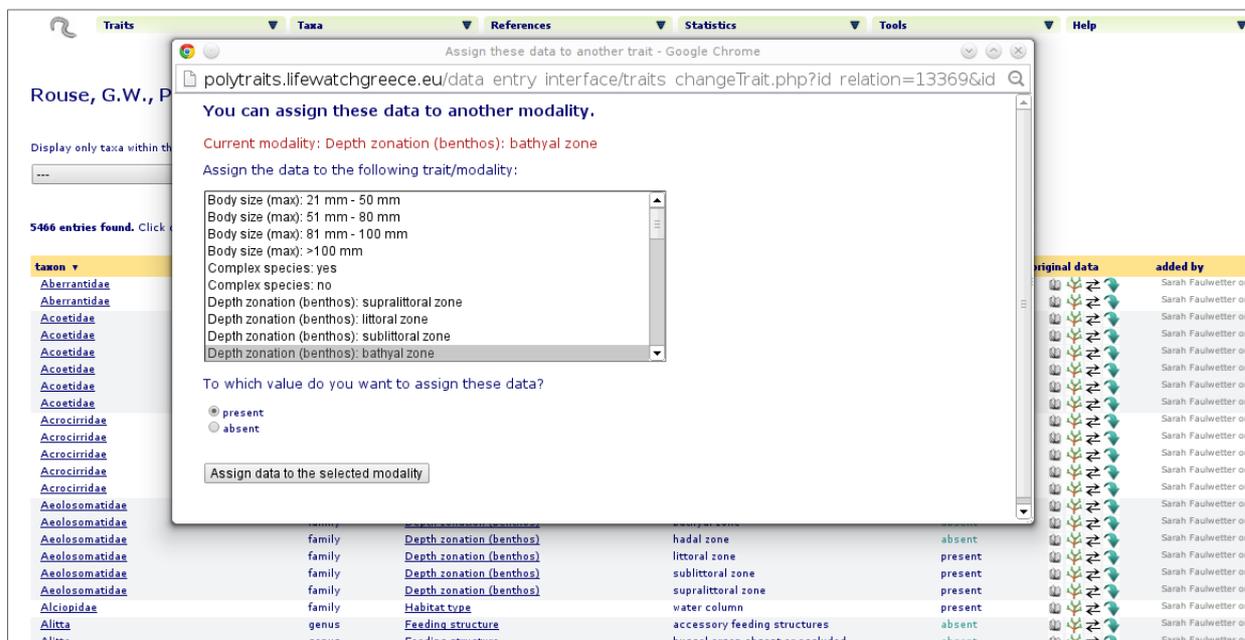


Figure 18: By clicking on the blue arrow, a dialogue opens allowing the user to assign the information to all or selected child taxa

trait database
export the data as a matrix

↶
Traits
▼
Taxa
▼
References
▼
Statistics
▼
Tools
▼
Help
▼

To export the data from the database as a cross table (a matrix, with species as rows and modalities as columns), choose the file format and if you want to infer missing information from higher levels as well as how to deal with missing information. Depending on the size of your dataset the inference might take up to a few minutes.

Taxon level:

Subjective synonyms:

treat subjective synonyms as separate species
 treat subjective synonyms as 'equal' to the valid name (lump all data)

Inference:

no inference
 infer missing data from congeners
 infer missing data from family members

Missing modalities:
Decide what to do when at least one modality (but not all) per trait is filled in:

don't do anything
 first infer data from other levels according to choice above, then fill in missing modalities as 'absent'
 first fill in missing modalities as 'absent', for remaining traits without any data infer data according to choice above. After each inference level missing modalities are set to 'absent'

Context (dataset):

Traits (by default, all traits will be exported):

- Body size (max)
- Complex species
- Depth zonation (benthos)
- Depth zonation (pelagic)
- Ecosystem engineering
- Feeding structure
- Feeding type
- Habitat type
- Intra- and interspecific competition
- Lifespan
- Migrations of adult
- Mobility of adult
- Physiographic feature
- Predated by
- Sociability

Filetype:

Export option 19: Data for the species including data from subjective synonyms are exported, then missing modalities of incomplete traits are set to 'absent', for remaining 'empty' traits data are first inferred from synonyms, then from congeners, then from other family members. Missing modalities of incomplete traits are set to 'absent' after each inference round.

Figure 19: Export interface for the data. In contrast to the public export interface, the internal export interface provides additional options, such as limiting the data to certain traits or taxa (within a context). The interface provides also a variety of options of how to treat synonyms (as separate entries or merged under the accepted name) and how to handle missing data (by inferring information from higher taxonomic levels or by setting missing information to “absent”).

Appendix III

Supplementary material of Chapter 5

Supplementary Material 1

Matrix indicating species' presence or absence in each of the lagoons

Taxon	AGI	GM	GVP	LOG	MdS	VAR
<i>Alitta succinea</i>						+
<i>Ampharete acutifrons</i>			+			
<i>Aricidea claudiae</i>						+
<i>Armandia cirrhosa</i>				+		
<i>Capitella capitata capitata</i>	+		+	+	+	
<i>Capitella minima</i>						+
<i>Eteone longa</i>				+		
<i>Eteone picta</i>		+	+			+
<i>Euclymene oerstedii</i>		+				
<i>Eunice vittata</i>		+				
<i>Fabricia stellaris stellaris</i>						+
<i>Ficopomatus enigmaticus</i>						+
<i>Glycera alba</i>		+				
<i>Glycera tridactyla</i>	+	+			+	+
<i>Glycera unicornis</i>		+				
<i>Harmothoe impar</i>						+
<i>Harmothoe spinifera</i>				+		
<i>Hediste diversicolor</i>		+	+		+	
<i>Heteromastus filiformis</i>	+	+	+	+		+
<i>Hydroides dianthus</i>	+			+		
<i>Leiochone leiopygos</i>	+					
<i>Malacoceros fuliginosus</i>		+	+		+	
<i>Malmgreniella lunulata</i>				+		
<i>Melinna palmata</i>						+
<i>Naineris laevigata</i>				+	+	
<i>Neanthes caudata</i>		+	+			
<i>Nephtys hombergii</i>		+	+	+		+
<i>Nereiphylla pusilla</i>				+		
<i>Nereiphylla rubiginosa</i>						+
<i>Nereis splendida</i>	+					
<i>Paradoneis lyra</i>	+					
<i>Pectinaria koreni</i>						+
<i>Phyllodoce mucosa</i>						+
<i>Platynereis dumerilii</i>						+
<i>Polydora ciliata</i>	+	+	+	+		+
<i>Prionospio caspersi</i>		+				
<i>Prionospio cirrifera</i>			+			+
<i>Prionospio fallax</i>				+		
<i>Protodorvillea kefersteini</i>						+
<i>Salvatoria clavata</i>						+
<i>Scolecopsis cantabra</i>				+		
<i>Scolecopsis tridentata</i>				+		
<i>Spio decoratus</i>	+	+				
<i>Spio filicornis</i>						+
<i>Sternaspis scutata</i>			+			
<i>Streblospio shrubsolii</i>		+	+			+
<i>Syllides fulvus</i>						+

Supplementary material 2

Species-by-trait matrix. Abbreviations of trait categories can be found in Supplementary material 3

Taxon / Trait	BS_2	BS_3	BS_4	BS_5	BS_6	BS_7	BS_1	CPLX	DZ_ABY	DZ_SUP	DZ_HAD	DZ_LIT	DZ_SUB	DZ_BAT	DEV_V	DEV_O
Alitta succinea	0	0	1	1	1	1	0	0	0	0	0	1	1	0	0	1
Ampharete acutifrons	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	1
Aricidea claudiae	1	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1
Armandia cirrhosa	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Capitella capitata capitata	0	0	0	0	1	0	0	1	0	0	0	1	1	1	0	1
Capitella minima	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1
Eteone longa	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	1
Eteone picta	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1
Euclymene oerstedii	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0	1
Eunice vittata	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1
Fabricia stellaris stellaris	1	0	0	0	0	0	1	0	0	0	0	1	1	1	0	1
Ficopomatus enigmaticus	0	0	1	1	0	0	0	0	0	0	0	1	1	0	0	1
Glycera alba	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	1
Glycera tridactyla	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1
Glycera unicornis	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0	1
Harmothoe impar	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	1
Harmothoe spinifera	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	1
Hediste diversicolor	0	0	0	0	0	1	0	1	0	1	0	1	1	0	1	1
Heteromastus filiformis	0	0	0	0	0	1	0	1	1	0	0	1	1	1	0	1
Hydroides dianthus	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1
Leiochone leiopygos	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	1
Malacoceros fuliginosus	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	1
Malmgreniella lunulata	0	1	1	0	0	0	0	0	0	0	0	1	1	1	0	1
Melinna palmata	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	1
Naineris laevigata	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	1
Neanthes caudata	0	0	1	0	0	0	0	1	0	0	0	1	1	1	0	1
Nephtys hombergii	0	0	0	0	0	1	0	1	0	0	0	1	1	1	0	1
Nereiphylla pusilla	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Nereiphylla rubiginosa	0	0	1	1	0	0	0	0	0	0	0	1	1	0	0	1
Nereis splendida	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1
Paradoneis lyra	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1
Pectinaria koreni	0	0	1	0	0	0	0	1	0	0	0	1	1	1	0	1
Perinereis cultrifera	0	0	0	0	0	1	0	1	0	1	0	1	1	0	0	1
Phyllodoce mucosa	0	0	0	0	0	1	0	0	0	0	0	1	1	1	0	1
Platynereis dumerilii	0	0	1	0	0	0	0	1	1	0	0	1	1	1	0	1
Polydora ciliata	0	0	1	0	0	0	0	1	0	0	0	1	1	1	0	1
Prionospio caspersi	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1
Prionospio cirrifer	0	1	0	0	0	0	0	0	1	0	0	1	1	1	0	1
Prionospio fallax	0	0	1	0	0	0	0	0	1	0	1	1	1	1	0	1
Protodorvillea kefersteini	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1
Salvatoria clavata	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1
Scolecopsis cantabra	0	0	1	0	1	0	0	0	0	0	0	1	1	0	0	1
Scolecopsis tridentata	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1
Spio decoratus	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1
Spio filicornis	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1
Sternaspis scutata	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1
Streblospio shrubsolii	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0
Syllides fulvus	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1

Taxon / Trait	EGG_L	EGG_M	EGG_S	FAC_LUN	FAC_HOR	FAC_PHO	FAC_SAL	FAC_TMP	FEC_100000	FEC_20000	FEC_500	FEC_2500	FEC_10000	FEC_100000+	FEC_50
Alitta succinea	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1
Ampharete acutifrons	0	1	0	0	0	0	0	1	0	1	0	1	0	0	0
Aricidea claudiae	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
Armandia cirrhosa	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0
Capitella capitata capitata	1	1	1	1	1	0	0	0	0	1	1	1	1	0	0
Capitella minima	1	1	1	1	1	0	0	0	0	0	1	0	0	0	1
Eteone longa	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0
Eteone picta	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0
Euclymene oerstedii	1	1	0	1	0	0	1	0	0	0	0	0	1	0	1
Eunice vittata	1	0	0	1	1	0	0	0	1	1	0	0	1	0	0
Fabricia stellaris stellaris	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0
Ficopomatus enigmaticus	0	0	1	0	0	0	1	1	0	0	0	1	1	0	0
Glycera alba	0	1	0	1	0	1	0	1	0	0	0	0	0	1	0
Glycera tridactyla	0	1	0	1	0	1	0	1	0	0	0	0	0	1	0
Glycera unicornis	0	1	0	1	0	1	0	1	0	0	0	0	0	1	0
Harmothoe impar	0	1	1	1	1	1	0	1	0	0	0	0	1	0	0
Harmothoe spinifera	0	0	1	1	1	1	0	1	0	0	0	0	1	0	0
Hediste diversicolor	1	0	0	1	1	1	0	1	0	0	0	1	1	0	0
Heteromastus filiformis	0	1	0	1	1	0	0	0	0	1	1	1	1	0	1
Hydroides dianthus	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0
Leiochone leiopygos	1	1	0	1	0	0	1	0	0	0	0	0	1	0	1
Malacoceros fuliginosus	0	1	0	0	1	1	1	1	0	1	0	0	0	0	0
Malmgreniella lunulata	0	0	1	1	1	1	0	1	0	0	0	0	1	1	0
Melinna palmata	1	1	0	0	0	0	0	1	0	0	0	1	0	0	0
Naineris laevigata	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0
Neanthes caudata	1	0	0	0	1	0	0	0	0	0	1	1	0	0	1
Nephtys hombergii	0	1	0	1	1	1	0	1	0	1	0	0	0	0	0
Nereiphylla pusilla	1	1	0	1	1	0	0	1	0	1	0	1	1	0	0
Nereiphylla rubiginosa	0	0	1	1	1	0	0	1	0	1	0	1	1	0	0
Nereis splendida	1	1	0	1	0	1	0	0	0	0	0	0	0	1	0
Paradoneis lyra	1	1	1	0	0	1	1	1	0	0	1	0	0	0	0
Pectinaria koreni	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0
Perinereis cultrifera	1	0	0	1	1	0	0	1	1	0	0	0	0	0	0
Phyllodoce mucosa	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0
Platynereis dumerilii	0	1	1	1	1	1	0	1	1	0	0	1	1	0	0
Polydora ciliata	0	1	1	0	1	0	0	1	0	0	1	1	1	0	0
Prionospio caspersi	0	1	1	0	0	0	1	1	0	0	0	1	0	0	0
Prionospio cirrifera	0	1	1	0	0	0	1	1	0	0	0	1	0	0	0
Prionospio fallax	0	0	1	0	0	0	1	1	0	0	0	1	0	0	0
Protodorvillea kefersteini	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1
Salvatoria clavata	0	1	1	0	1	0	0	0	0	0	1	0	0	0	1
Scolecopsis cantabra	0	0	1	0	1	1	1	1	0	0	1	0	0	0	0
Scolecopsis tridentata	0	1	0	0	1	1	1	1	0	0	1	0	0	0	0
Spio decoratus	0	1	1	0	1	1	1	1	0	0	1	1	0	0	0
Spio filicornis	1	1	1	0	1	1	1	1	0	0	1	0	0	0	0
Sternaspis scutata	1	1	0	0	1	0	0	0	0	0	1	1	1	0	0
Streblospio shrubsolii	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1
Syllides fulvus	0	1	1	1	1	1	0	1	0	0	0	0	0	0	1

Taxon / Trait	FEED_C	FEED_F	FEED_O	FEED_S	FEED_H	FEED_D	PER_EXT	PER_INT	MAT_4Y	MAT_2M	MAT_1Y	MAT_2Y	MAT_3Y	MAT_6M	LDEV_D	LDEV_I
<i>Alitta succinea</i>	1	0	1	0	0	1	1	0	0	1	0	0	0	0	0	1
<i>Ampharete acutifrons</i>	0	0	0	0	0	1	1	0	0	0	1	1	0	0	1	0
<i>Aricidea claudiae</i>	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0
<i>Armandia cirrhosa</i>	0	0	0	0	0	1	1	0	0	1	1	0	0	1	0	1
<i>Capitella capitata capitata</i>	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1
<i>Capitella minima</i>	0	0	0	1	1	1	1	0	0	1	0	0	0	1	0	1
<i>Eteone longa</i>	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1
<i>Eteone picta</i>	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1
<i>Euclymene oerstedii</i>	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	1
<i>Eunice vittata</i>	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1
<i>Fabricia stellaris stellaris</i>	0	1	0	0	0	0	1	0	0	0	0	1	1	0	1	0
<i>Ficopomatus enigmaticus</i>	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1
<i>Glycera alba</i>	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1
<i>Glycera tridactyla</i>	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1
<i>Glycera unicornis</i>	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1
<i>Harmothoe impar</i>	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1
<i>Harmothoe spinifera</i>	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1
<i>Hediste diversicolor</i>	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1
<i>Heteromastus filiformis</i>	0	0	0	0	1	1	1	1	0	0	0	1	0	0	0	1
<i>Hydroides dianthus</i>	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1
<i>Leiochone leiopygos</i>	0	0	0	0	0	1	1	0	0	0	0	1	1	0	1	1
<i>Malacoceros fuliginosus</i>	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1
<i>Malmgreniella lunulata</i>	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1
<i>Melinna palmata</i>	0	1	0	0	0	1	1	0	0	0	0	1	1	0	0	1
<i>Naineris laevigata</i>	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	1
<i>Neanthes caudata</i>	1	0	1	0	1	1	1	0	0	1	0	0	0	1	1	0
<i>Nephtys hombergii</i>	1	0	0	1	0	0	1	0	0	0	0	1	1	0	0	1
<i>Nereiphylla pusilla</i>	1	0	0	0	0	0	1	0	0	0	1	1	1	0	0	1
<i>Nereiphylla rubiginosa</i>	1	0	0	0	0	0	1	0	0	0	1	1	1	0	0	1
<i>Nereis splendida</i>	0	0	1	0	1	0	1	0	0	0	1	0	0	0	0	1
<i>Paradoneis lyra</i>	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0
<i>Pectinaria koreni</i>	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
<i>Perinereis cultrifera</i>	1	0	1	0	1	0	1	0	0	0	0	1	1	0	1	1
<i>Phyllodoce mucosa</i>	1	0	0	1	0	0	1	0	0	0	1	1	1	0	0	1
<i>Platynereis dumerilii</i>	0	0	1	0	1	0	1	0	0	0	1	1	0	1	1	1
<i>Polydora ciliata</i>	1	1	0	1	0	1	1	0	0	1	1	0	0	0	1	1
<i>Prionospio caspersi</i>	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
<i>Prionospio cirrifera</i>	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	1
<i>Prionospio fallax</i>	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	1
<i>Protodorvillea kefersteini</i>	1	0	0	1	1	0	1	0	0	0	0	0	0	1	1	0
<i>Salvatoria clavata</i>	1	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0
<i>Scolecopsis cantabra</i>	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1
<i>Scolecopsis tridentata</i>	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0
<i>Spio decoratus</i>	0	1	0	0	0	1	1	0	0	1	0	0	0	0	1	1
<i>Spio filicornis</i>	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0
<i>Sternaspis scutata</i>	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	1
<i>Streblospio shrubsolii</i>	0	1	0	0	0	1	0	1	0	1	0	0	0	0	1	0
<i>Syllides fulvus</i>	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	1

Taxon / Trait	LFT_M	LFT_P	LM_B	LM_P	LIFE_3-5	LIFE_1-3	LIFE_5+	LIFE_1	PC_NEAR	PC_FAR	TUBE_GRAY	TUBE_CLAY	TUBE_SECR	TUBE_SAND	TUBE_CALC	TUBE_MUD	TUBE_BIO
<i>Alitta succinea</i>	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0
<i>Ampharete acutifrons</i>	1	0	1	0	0	1	0	1	1	0	0	0	1	1	0	1	0
<i>Aricidea claudiae</i>	1	0	1	1	0	0	0	1	1	0	0	0	1	1	0	1	0
<i>Armandia cirrhosa</i>	0	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0
<i>Capitella capitata capitata</i>	1	0	1	1	0	1	0	0	1	0	0	0	1	1	0	0	1
<i>Capitella minima</i>	1	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0
<i>Eteone longa</i>	0	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0
<i>Eteone picta</i>	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0
<i>Euclymene oerstedii</i>	1	0	1	1	1	0	1	0	0	0	0	0	1	1	0	0	0
<i>Eunice vittata</i>	1	1	0	1	0	1	1	0	0	1	0	0	1	1	0	0	1
<i>Fabricia stellaris stellaris</i>	1	0	1	0	0	1	0	0	1	0	0	0	1	1	0	1	1
<i>Ficopomatus enigmaticus</i>	0	1	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0
<i>Glycera alba</i>	0	1	1	1	0	0	1	0	0	0	0	0	1	0	0	0	1
<i>Glycera tridactyla</i>	0	1	0	1	1	0	1	0	0	0	0	0	1	1	0	0	1
<i>Glycera unicornis</i>	0	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0
<i>Harmothoe impar</i>	0	1	0	1	1	1	0	0	1	0	0	0	0	0	1	0	0
<i>Harmothoe spinifera</i>	0	1	0	1	1	1	0	0	1	0	0	0	0	0	1	0	0
<i>Hediste diversicolor</i>	1	0	1	1	0	1	0	0	1	0	0	0	1	0	0	0	0
<i>Heteromastus filiformis</i>	1	1	1	1	1	1	0	0	0	1	0	0	1	0	0	1	0
<i>Hydroides dianthus</i>	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
<i>Leiochone leiopygos</i>	1	0	1	1	1	0	1	0	0	0	0	0	1	1	0	0	1
<i>Malacoceros fuliginosus</i>	0	1	0	1	0	1	0	0	0	1	1	0	1	0	0	0	0
<i>Malmgreniella lunulata</i>	0	1	0	1	1	1	0	0	0	0	0	0	0	0	1	0	0
<i>Melinna palmata</i>	1	0	1	1	0	1	1	0	0	0	0	0	1	0	0	1	1
<i>Naineris laevigata</i>	1	0	1	1	1	0	0	0	1	0	0	0	0	0	1	0	0
<i>Neanthes caudata</i>	1	1	1	0	0	0	0	1	1	0	0	0	1	1	0	1	1
<i>Nephtys hombergii</i>	1	1	1	1	0	1	1	0	0	0	0	0	1	0	0	0	0
<i>Nereiphylla pusilla</i>	1	0	0	1	0	1	0	0	0	1	0	0	1	0	0	0	0
<i>Nereiphylla rubiginosa</i>	1	1	0	1	0	1	0	0	0	1	0	0	1	0	0	0	0
<i>Nereis splendida</i>	1	0	0	1	0	1	0	0	1	1	1	0	1	1	0	1	1
<i>Paradoneis lyra</i>	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0
<i>Pectinaria koreni</i>	0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	0	0
<i>Perinereis cultrifera</i>	1	1	0	1	1	0	0	0	0	1	1	0	0	1	0	1	0
<i>Phyllodoce mucosa</i>	0	1	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0
<i>Platynereis dumerilii</i>	1	1	1	1	0	1	0	1	0	0	0	0	1	1	0	0	1
<i>Polydora ciliata</i>	0	1	0	1	0	1	0	1	1	0	0	0	1	1	0	1	1
<i>Prionospio caspersi</i>	0	1	0	1	0	1	0	1	0	0	0	0	1	1	0	1	0
<i>Prionospio cirrifera</i>	1	1	0	1	0	1	0	0	0	0	0	0	1	1	0	1	0
<i>Prionospio fallax</i>	0	1	1	1	0	1	0	1	0	0	0	0	1	0	0	0	0
<i>Protodorvillea kefersteini</i>	1	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0
<i>Salvatoria clavata</i>	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1	0	0
<i>Scolecopsis cantabra</i>	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
<i>Scolecopsis tridentata</i>	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0
<i>Spio decoratus</i>	1	1	1	1	0	0	0	1	0	1	0	0	1	1	0	0	0
<i>Spio filicornis</i>	0	1	1	1	0	1	0	1	1	0	0	0	1	1	0	0	0
<i>Sternaspis scutata</i>	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
<i>Streblospio shrubsolii</i>	1	1	1	1	0	0	0	1	1	0	0	0	1	1	0	1	0
<i>Syllides fulvus</i>	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0

Taxon / Trait	MOB_BUR	MOB_SWIM	MOB_CRAWL	MOB_SESS	REP_GON	REP_SEQ	REP_SIM	PC_YES	PC_NO	OOG_INTRA	OOG_EXTRA	STRAT_SEM	STRAT_ITER
Alitta succinea	1	1	1	0	1	0	0	0	1	0	1	1	0
Ampharete acutifrons	0	1	0	1	1	0	0	1	1	0	1	1	1
Aricidea claudiae	1	1	0	0	1	0	0	1	0	1	0	1	0
Armandia cirrhosa	1	1	0	0	1	0	0	0	1	0	1	0	1
Capitella capitata capitata	1	0	0	0	1	1	0	1	0	1	0	0	1
Capitella minima	1	0	0	0	0	0	1	1	0	1	0	0	1
Eteone longa	1	1	1	0	1	0	0	0	1	0	1	1	0
Eteone picta	1	1	1	0	1	0	0	0	1	0	1	1	0
Euclymene oerstedii	0	0	0	1	1	0	1	0	1	0	1	0	1
Eunice vittata	0	1	1	0	1	0	0	1	1	0	1	0	1
Fabricia stellaris stellaris	0	1	1	1	1	0	0	1	0	0	1	0	1
Ficopomatus enigmaticus	0	0	0	1	1	1	1	0	1	0	1	0	1
Glycera alba	1	1	0	0	1	0	0	0	1	0	1	1	0
Glycera tridactyla	1	1	0	0	1	0	0	0	1	0	1	1	0
Glycera unicornis	1	1	0	0	1	0	0	0	1	0	1	1	0
Harmothoe impar	0	0	1	0	1	0	0	1	0	1	0	0	1
Harmothoe spinifera	0	0	1	0	1	0	0	1	1	1	0	1	1
Hediste diversicolor	1	1	1	0	1	0	0	1	0	0	1	1	0
Heteromastus filiformis	1	0	0	0	1	0	0	1	1	1	0	1	0
Hydroides dianthus	0	0	0	1	1	0	0	0	1	0	1	1	1
Leiochone leiopygos	0	0	0	1	1	0	0	0	1	0	1	0	1
Malacoceros fuliginosus	1	0	0	1	1	0	0	1	1	0	1	0	1
Malmgreniella lunulata	0	0	1	0	1	0	0	0	1	1	0	1	1
Melinna palmata	0	0	0	1	1	0	0	0	1	0	1	0	1
Naineris laevigata	1	0	0	0	1	1	0	1	0	1	0	0	1
Neanthes caudata	1	0	1	0	1	0	0	1	0	0	1	1	1
Nephtys hombergii	1	1	1	0	1	0	0	0	1	1	0	0	1
Nereiphylla pusilla	0	0	1	0	1	0	0	1	1	0	1	1	1
Nereiphylla rubiginosa	0	0	1	0	1	0	0	1	1	0	1	1	1
Nereis splendida	1	1	1	0	1	0	0	1	1	0	1	1	0
Paradoneis lyra	1	1	0	0	1	0	0	1	0	1	0	1	0
Pectinaria koreni	1	0	0	1	1	0	1	0	1	0	1	1	0
Perinereis cultrifera	1	1	1	0	1	0	0	1	1	0	1	1	0
Phyllodoce mucosa	1	1	1	0	1	0	0	1	0	0	1	1	1
Platynereis dumerilii	0	1	1	1	1	0	0	0	1	0	1	1	0
Polydora ciliata	1	0	0	1	1	0	0	1	0	0	1	1	1
Prionospio caspersi	1	0	0	1	1	0	0	0	1	1	0	1	0
Prionospio cirrifera	0	0	0	1	1	0	0	0	1	1	0	1	0
Prionospio fallax	1	1	0	1	1	0	0	0	1	1	0	1	0
Protodorvillea kefersteini	0	1	1	0	1	0	0	0	1	0	1	1	0
Salvatoria clavata	0	1	1	0	0	1	0	1	0	0	1	0	1
Scolecopsis cantabra	1	0	0	1	1	0	0	0	1	1	0	1	1
Scolecopsis tridentata	1	0	0	1	1	0	0	1	0	1	0	1	1
Spio decoratus	0	0	0	1	1	0	0	1	0	0	1	1	0
Spio filicornis	1	0	0	1	1	0	0	1	0	0	1	1	0
Sternaspis scutata	1	0	0	0	1	0	0	0	1	1	0	0	1
Streblospio shrubsolii	0	0	0	1	1	0	0	1	0	1	0	0	1
Syllides fulvus	0	1	1	0	1	0	0	1	0	1	1	1	0

Taxon / Trait	RW_OEV	RW_BEV	RW_REG	RW_DC	RW_UC	RW_DIFH	RW_HAB	SM_NO	SM_YES	FREQ_CONT	FREQ_ANN	FREQ_MULT	SPERM_ENT	SPERM_ECT	SPERM_INTRO
Alitta succinea	1	0	0	0	1	1	0	0	1	0	1	0	0	1	0
Ampharete acutifrons	0	0	0	0	0	1	0	1	0	0	1	1	0	1	1
Aricidea claudiae	0	0	0	0	1	0	0	0	1	1	0	0	1	0	0
Armandia cirrhosa	0	0	0	0	0	1	0	1	1	1	0	0	0	1	0
Capitella capitata capitata	0	0	0	1	0	0	0	1	0	1	1	0	0	0	1
Capitella minima	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1
Eteone longa	0	0	1	0	0	1	0	0	1	1	1	0	0	1	0
Eteone picta	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0
Euclymene oerstedii	1	1	0	0	1	1	0	1	0	0	1	0	1	0	0
Eunice vittata	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0
Fabricia stellaris stellaris	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0
Ficopomatus enigmaticus	0	0	0	0	0	0	1	0	1	1	1	0	0	1	0
Glycera alba	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0
Glycera tridactyla	0	0	0	0	0	1	0	0	1	0	1	1	0	1	0
Glycera unicornis	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0
Harmothoe impar	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0
Harmothoe spinifera	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0
Hediste diversicolor	1	0	1	1	0	1	0	1	1	1	1	0	0	1	0
Heteromastus filiformis	0	1	0	0	1	1	0	1	0	0	1	0	0	1	0
Hydroides dianthus	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0
Leiochone leiopygos	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0
Malacoceros fuliginosus	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0
Malmgreniella lunulata	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0
Melinna palmata	0	0	0	0	0	1	0	1	0	1	1	0	0	1	0
Naineris laevigata	0	0	0	0	1	0	0	1	0	1	1	0	1	0	0
Neanthes caudata	0	0	0	0	0	1	0	1	0	1	0	0	0	1	0
Nephtys hombergii	0	0	1	0	0	1	0	0	1	1	1	0	0	1	0
Nereiphylla pusilla	0	0	1	0	0	1	0	0	1	1	1	0	0	1	0
Nereiphylla rubiginosa	0	0	1	0	0	1	0	0	1	1	1	0	0	1	0
Nereis splendida	1	0	1	1	1	1	0	1	1	0	1	0	0	1	0
Paradoneis lyra	0	0	0	0	1	1	0	0	1	0	1	0	1	0	0
Pectinaria koreni	1	0	0	0	1	0	0	1	0	0	1	1	0	1	0
Perinereis cultrifera	1	0	0	0	0	1	0	1	1	1	0	0	0	1	0
Phyllodoce mucosa	0	0	0	0	0	1	0	1	0	0	1	0	0	1	0
Platynereis dumerilii	0	0	0	0	0	1	0	0	1	1	0	0	0	1	0
Polydora ciliata	0	0	0	1	1	1	0	1	0	1	1	1	0	0	1
Prionospio caspersi	0	0	0	1	1	1	0	1	0	0	1	0	0	1	0
Prionospio cirrifera	0	0	0	0	1	1	0	1	0	0	1	0	0	1	0
Prionospio fallax	0	0	0	0	0	1	0	1	0	0	1	0	0	1	0
Protodorvillea kefersteini	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0
Salvatoria clavata	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0
Scolecopsis cantabra	0	0	0	0	1	1	0	1	0	0	1	0	0	1	0
Scolecopsis tridentata	0	0	0	0	1	1	0	1	0	1	1	0	0	1	0
Spio decoratus	0	0	0	0	0	1	0	1	0	1	1	0	0	0	1
Spio filicornis	0	0	0	1	1	0	0	1	0	0	1	0	0	0	1
Sternaspis scutata	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0
Streblospio shrubsolii	0	0	1	1	0	0	0	1	0	1	0	0	0	0	1
Syllides fulvus	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0

Taxon / Trait	SAL_FULLL	SAL_RCD	SAL_LOW	TEMP_WARM	TEMP_HOT	TEMP_COLD	EPKY_NO	EPKY_YES	SYNC_NO	SYNC_YES	TOL_IV	TOL_V	TOL_I	TOL_II	TOL_III
Alitta succinea	1	1	1	1	1	1	0	1	0	1	0	0	0	0	1
Ampharete acutifrons	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0
Aricidea claudiae	1	1	1	1	0	1	0	1	1	0	0	0	0	0	1
Armandia cirrhosa	1	1	1	1	1	0	0	1	1	0	0	0	1	0	0
Capitella capitata capitata	1	1	1	1	1	1	1	0	1	0	0	1	0	0	0
Capitella minima	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0
Eteone longa	1	1	1	1	1	1	0	1	0	1	0	0	0	1	0
Eteone picta	1	1	1	1	1	1	0	1	0	1	0	0	0	1	0
Euclymene oerstedii	1	0	0	1	0	1	1	0	0	1	0	0	1	0	0
Eunice vittata	1	0	0	1	1	0	0	1	0	1	0	0	0	1	0
Fabricia stellaris stellaris	1	1	1	1	0	1	1	0	1	0	0	0	0	1	0
Ficopomatus enigmaticus	1	1	1	1	1	1	0	1	0	1	0	0	0	0	1
Glycera alba	1	1	1	1	1	1	0	1	0	1	0	0	1	0	0
Glycera tridactyla	1	0	0	1	1	0	0	1	0	1	0	0	1	0	0
Glycera unicornis	1	1	0	1	1	0	0	1	0	1	0	0	1	0	0
Harmothoe impar	1	1	1	1	0	1	1	0	0	1	0	0	0	1	0
Harmothoe spinifera	1	1	0	1	0	0	1	0	0	1	0	0	0	1	0
Hediste diversicolor	1	1	1	1	1	1	1	0	1	1	0	0	0	0	1
Heteromastus filiformis	1	1	1	1	1	1	1	0	1	0	1	0	0	0	1
Hydroides dianthus	1	1	0	1	1	1	0	1	0	1	0	0	0	0	1
Leiochone leiopygos	1	0	0	1	0	0	1	0	0	1	0	0	0	0	1
Malacoceros fuliginosus	1	1	0	1	1	0	1	0	1	0	0	1	0	0	0
Malmgreniella lunulata	1	1	1	1	1	1	1	0	1	1	0	0	0	1	0
Melinna palmata	1	0	0	1	1	1	1	0	0	1	0	0	0	0	1
Naineris laevigata	0	1	0	1	1	1	1	0	1	1	0	0	1	0	0
Neanthes caudata	1	1	1	1	0	0	1	0	1	0	0	0	0	0	1
Nephtys hombergii	1	1	1	1	0	1	1	0	0	1	0	0	0	1	0
Nereiphylla pusilla	1	1	1	1	1	1	0	1	0	1	0	0	0	1	0
Nereiphylla rubiginosa	1	1	1	1	1	1	0	1	0	1	0	0	0	1	0
Nereis splendida	1	1	0	1	1	1	1	0	1	0	0	0	0	1	0
Paradoneis lyra	1	0	0	1	0	1	0	1	1	0	0	0	0	0	1
Pectinaria koreni	1	1	1	1	0	1	1	0	1	0	0	0	1	0	0
Perinereis cultrifera	1	1	0	1	1	0	1	1	1	1	0	0	0	0	1
Phyllodoce mucosa	1	1	1	1	1	1	1	0	0	1	0	0	0	0	1
Platynereis dumerilii	1	1	1	1	1	0	0	1	0	1	0	0	0	0	1
Polydora ciliata	1	1	1	1	1	1	1	0	1	0	1	0	0	0	0
Prionospio caspersi	1	0	0	1	0	1	1	0	1	0	1	0	0	0	0
Prionospio cirrifer	1	0	0	1	0	1	1	0	0	1	1	0	0	0	0
Prionospio fallax	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0
Protodorvillea kefersteini	1	1	0	1	0	1	0	1	1	0	0	0	0	1	0
Salvatoria clavata	1	1	1	1	1	1	0	1	0	1	0	0	0	1	1
Scolecopsis cantabra	1	0	0	1	0	1	1	0	1	0	0	0	0	1	0
Scolecopsis tridentata	1	1	0	1	0	1	1	0	1	0	0	0	0	1	0
Spio decoratus	1	1	1	1	1	1	1	0	1	0	0	0	0	0	1
Spio filicornis	1	1	1	1	1	1	1	0	1	0	0	0	0	0	1
Sternaspis scutata	1	1	1	1	0	1	1	0	1	0	0	0	0	0	1
Streblospio shrubsolii	1	1	1	1	1	1	1	0	1	0	0	0	0	0	1
Syllides fulvus	1	0	0	1	1	1	0	1	1	1	0	0	0	1	0

Taxon / Trait	STRUCT_VMP	STRUCT_YBO	STRUCT_SAP	STRUCT_MAP	STRUCT_ACC	STRUCT_ABS	SUBST_CLAY	SUBST_HARD	SUBST_GRAY	SUBST_SAND	SUBST_MUD
Alitta succinea	0	0	0	1	0	0	1	1	1	1	1
Ampharete acutifrons	0	1	0	0	1	0	1	1	1	1	1
Aricidea claudiae	0	0	1	0	0	0	0	1	0	0	1
Armandia cirrhosa	0	0	1	0	0	0	0	0	1	1	1
Capitella capitata capitata	0	0	1	0	0	0	0	1	1	1	1
Capitella minima	0	0	1	0	0	0	0	1	1	1	1
Eteone longa	0	0	0	1	0	0	0	1	1	1	1
Eteone picta	0	0	0	1	0	0	0	1	1	1	1
Euclymene oerstedii	0	0	1	0	0	0	1	1	0	1	1
Eunice vittata	1	0	0	0	0	0	1	1	1	1	1
Fabricia stellaris stellaris	0	0	0	0	1	1	0	1	0	1	1
Ficopomatus enigmaticus	0	0	0	0	1	1	0	1	0	1	0
Glycera alba	0	0	0	1	0	0	0	1	1	1	1
Glycera tridactyla	0	0	0	1	0	0	0	1	0	1	1
Glycera unicornis	0	0	0	1	0	0	0	0	0	1	1
Harmothoe impar	0	0	0	1	0	0	0	1	0	0	1
Harmothoe spinifera	0	0	0	1	0	0	0	1	0	0	0
Hediste diversicolor	0	0	0	1	0	0	1	1	1	1	1
Heteromastus filiformis	0	0	1	0	0	0	0	1	0	1	1
Hydroides dianthus	0	0	0	0	1	1	0	1	0	0	0
Leiochone leiopygos	0	0	1	0	0	0	0	0	0	1	1
Malacoceros fuliginosus	0	0	1	0	1	0	0	0	1	1	1
Malmgreniella lunulata	0	0	0	1	0	0	0	1	1	1	1
Melinna palmata	0	1	0	0	1	0	1	0	1	1	1
Naineris laevigata	0	0	1	0	0	0	0	1	0	1	1
Neanthes caudata	0	0	0	1	0	0	1	1	1	1	1
Nephtys hombergii	0	0	0	1	0	0	0	0	0	1	1
Nereiphylla pusilla	0	0	0	1	0	0	0	1	0	1	0
Nereiphylla rubiginosa	0	0	0	1	0	0	0	1	0	1	1
Nereis splendida	0	0	0	1	0	0	0	1	0	1	0
Paradoneis lyra	0	0	1	0	0	0	0	1	1	1	1
Pectinaria koreni	0	1	0	0	1	0	1	1	1	1	1
Perinereis cultrifera	0	0	0	1	0	0	0	1	1	1	1
Phyllodoce mucosa	0	0	0	1	0	0	0	1	1	1	1
Platynereis dumerilii	0	0	0	1	0	0	0	1	1	1	1
Polydora ciliata	0	1	0	0	1	0	1	1	0	1	1
Prionospio caspersi	0	1	0	0	1	0	0	0	0	1	1
Prionospio cirrifera	0	1	0	0	1	0	1	0	0	1	1
Prionospio fallax	0	1	0	0	1	0	0	0	1	1	1
Protodorvillea kefersteini	1	0	0	0	0	0	0	1	0	1	1
Salvatoria clavata	0	0	0	1	0	0	0	1	1	1	1
Scolelepis cantabra	0	0	1	0	1	0	0	0	0	1	1
Scolelepis tridentata	0	0	1	0	1	0	0	0	0	1	1
Spio decoratus	0	1	0	0	1	0	0	1	0	1	1
Spio filicornis	0	1	0	0	1	0	0	0	1	1	1
Sternaspis scutata	0	0	1	0	0	0	1	1	1	1	1
Streblospio shrubsoleii	0	0	1	0	1	0	0	1	0	1	1
Syllides fulvus	0	0	0	1	0	0	0	1	0	1	0

Supplementary material 3

List of all modalities (and their abbreviations) used in the analysis. Full definitions of the terms can be found at <http://polytraits.lifewatchgreece.eu/terms>

Modality	Code
Age at first reproduction : < 2 months	MAT_2M
Age at first reproduction : 1–2 years	MAT_2Y
Age at first reproduction : 2–3 years	MAT_3Y
Age at first reproduction : 2–6 months	MAT_6M
Age at first reproduction : 3–4 years	MAT_4Y
Age at first reproduction : 6 months–1year	MAT_1Y
Body size (max) : >10cm (100mm)	BS_7
Body size (max) : 0.25cm -1 cm (2.5mm–10 mm)	BS_2
Body size (max) : 0.2mm - 0.25 cm (200µm–2.5 mm)	BS_1
Body size (max) : 1cm-2cm (10mm–20mm)	BS_3
Body size (max) : 2cm-5cm (20mm–50mm)	BS_4
Body size (max) : 5cm -8 cm (50mm–80mm)	BS_5
Body size (max) : 8cm -10cm (80mm–100mm)	BS_6
Complex species reported : yes	CPLX
Depth zonation (benthos) : supralittoral zone	DZ_SUP
Depth zonation (benthos) : littoral zone	DZ_LIT
Depth zonation (benthos) : sublittoral zone (down to 200m)	DZ_SUB
Depth zonation (benthos) : bathyal zone (200–4000m)	DZ_BAT
Depth zonation (benthos) : abyssal zone (4000–6000m)	DZ_ABY
Depth zonation (benthos) : hadal zone (below 6000m)	DZ_HAD
Developmental mechanism : Oviparous	DEV_O
Developmental mechanism : Viviparous	DEV_V
Ecosystem engineering : Biodiffusor	RW_DIFF
Ecosystem engineering : Blind-ended ventilation	RW_BEV
Ecosystem engineering : Downward conveyour	RW_DC
Ecosystem engineering : habitat-building (reef-forming)	RW_HAB
Ecosystem engineering : Open-ended ventilation	RW_OEV
Ecosystem engineering : Regenerator	RW_REG
Ecosystem engineering : Upward conveyor	RW_UC
Egg size : 0µm–100µm	EGG_S
Egg size : 100µm–200µm	EGG_M
Egg size : >200µm	EGG_L
Epitoky : no	EPKY_NO
Epitoky : yes	EPKY_YES
Factors triggering reproduction : Lunar cycle	FAC_LUN
Factors triggering reproduction : Pheromones/ hormones	FAC_HOR
Factors triggering reproduction : Photoperiod	FAC_PHO
Factors triggering reproduction : Salinity	FAC_SAL
Factors triggering reproduction : Temperature	FAC_TMP
Fecundity : 1–50	FEC_50
Fecundity : 50–500	FEC_500
Fecundity : 500–2500	FEC_2500
Fecundity : 2500–10.000	FEC_10000
Fecundity : 10.000–20.000	FEC_20000
Fecundity : 20.000–100.000	FEC_100000

Fecundity : >100000	FEC_100000+
Feeding structure : Accessory feeding structures	STRUCT_ACC
Feeding structure : Buccal organ absent or occluded	STRUCT_ABS
Feeding structure : Muscular axial proboscis	STRUCT_MAP
Feeding structure : Simple axial proboscis	STRUCT_SAP
Feeding structure : Ventral buccal organ (simple)	STRUCT_VBO
Feeding structure : Ventral muscular proboscis	STRUCT_VMP
Feeding type : Carnivore	FEED_C
Feeding type : Deposit feeder	FEED_D
Feeding type : Filter feeder	FEED_F
Feeding type : Herbivore	FEED_H
Feeding type : Omnivore	FEED_O
Feeding type : Scavenger	FEED_S
Fertilization : Internal	FER_INT
Fertilization : External	FER_EXT
Larval development : Direct	LDEV_D
Larval development : Indirect	LDEV_I
Larval feeding type : Maternally derived nutrition	LFT_M
Larval feeding type : Planktotrophic	LFT_P
Larval mode : benthic	LM_B
Larval mode : pelagic	LM_P
Lifespan/ longevity : 1 year	LIFE_1
Lifespan/ longevity : 1-3 years	LIFE_1-3
Lifespan/ longevity : 3-5 years	LIFE_3-5
Lifespan/ longevity : 5 years	LIFE_5+
Location of parental care : Outside microenvironment of the worm	PC_FAR
Location of parental care : Within microenvironment of the worm	PC_NEAR
Mobility of adult : burrower (motile, by peristalsis)	MOB_BUR
Mobility of adult : crawler	MOB_CRAWL
Mobility of adult : non-motile / semi-motile	MOB_SESS
Mobility of adult : swimmer	MOB_SWIM
Mode of reproduction : Gonochoristic	REP_GON
Mode of reproduction : Sequentially hermaphrodite	REP_SEQ
Mode of reproduction : Simultaneous hermaphrodite	REP_SIM
Parental care : no	BP_NO
Parental care : yes	BP_YES
Pattern of oogenesis : extraovarian	OOG_EXTRA
Pattern of oogenesis : intraovarian	OOG_INTRA
Reproduction strategy of the individual : Iteroparous	STRAT_ITER
Reproduction strategy of the individual : Semelparous	STRAT_SEM
Sexual metamorphosis : no	SM_NO
Sexual metamorphosis : yes	SM_YES
Spawning frequency of the population : Annually, seasonal	FREQ_ANN
Spawning frequency of the population : Continuous or semi-continuous	FREQ_CONT
Spawning frequency of the population : multiple events/year	FREQ_MULTI
Sperm type : Ect-aquasperm	SPERM_ECT
Sperm type : Ent-aquasperm	SPERM_ENT
Sperm type : Introsperm	SPERM_INTRO
Substrate type : Clay	SUBST_CLAY
Substrate type : Hard	SUBST_HARD
Substrate type : Gravel	SUBST_GRAV
Substrate type : Sand	SUBST_SAND
Substrate type : Mud	SUBST_MUD
Survival salinity : Low salinity <18	SAL_LOW

Appendix III

Survival salinity : Reduced salinity (18–30)	SAL_RCD
Survival salinity : Full salinity (30–40)	SAL_FULL
Survival temperature : Cold waters (<0–10°C)	TEMP_COLD
Survival temperature : Warm/ temperate/ subtropical waters (10–25°C)	TEMP_WARM
Survival temperature : Tropical waters (>25°C)	TEMP_HOT
Synchronization of spawning of the population : no	SYNC_NO
Synchronization of spawning of the population : yes	SYNC_YES
Tolerance (AMBI index) : Group I	TOL_I
Tolerance (AMBI index) : Group II	TOL_II
Tolerance (AMBI index) : Group II	TOL_III
Tolerance (AMBI index) : Group II	TOL_IV
Tolerance (AMBI index) : Group II	TOL_V
Tube/ burrow material : Secretions (usually mucus)	TUBE_SECR
Tube/ burrow material : Biogenic products	TUBE_BIO
Tube/ burrow material : Calcium carbonate	TUBE_CALC
Tube/ burrow material : Clay	TUBE_CLAY
Tube/ burrow material : Gravel	TUBE_GRAV
Tube/ burrow material : Mud	TUBE_MUD
Tube/ burrow material : Sand	TUBE_SAND

