Review

Cryptic Clitellata: Molecular Species Delimitation of Clitellate Worms (Annelida): An Overview

Svante Martinsson * and Christer Erséus

Systematics and Biodiversity, Department of Biological and Environmental Sciences, University of Gothenburg, Box 463, SE-405 30 Göteborg, Sweden; christer.erseus@bioenv.gu.se

* Correspondence: svante.martinsson@bioenv.gu.se

Abstract: Methods for species delimitation using molecular data have developed greatly and have become a staple in systematic studies of clitellate worms. Here we give a historical overview of the data and methods used to delimit clitellates from the mid-1970s to today. We also discuss the taxonomical treatment of the cryptic species, including the recommendation that cryptic species, as far as possible, should be described and named. Finally, we discuss the prospects and further development of the field.

Keywords: DNA-barcoding; Oligochaeta; cryptic species

1. Introduction

Species delimitation, i.e., the process of determining species boundaries and discovering species, is a field that has developed quickly since the introduction of genetic data [1,2]. The development has been both on the data side, from protein patterns to large genomic datasets, and on the analytical side, from clustering and measures of genetic distances to complex analyses based on coalescent theory. These advances have led to an increase in the discovery of cryptic species, i.e., species that are morphologically similar and, therefore, have been classified as the same nominal species [3]. Cryptic species are found all over the animal kingdom (e.g., [4,5]), including annelids (e.g., [6,7]) and, despite morphological similarities, they may differ in ecologically and physiologically important aspects (see, e.g., [8,9]). Species are basic biological units and entities of generalisation, and, therefore, the basis of most studies. A number of clitellate species are used as models in several fields, e.g., ecotoxicology, neurobiology and soil ecology [10,11] and, in several of the species used, taxonomical problems have been found [12–16]. In this kind of work, it is important to know the true identity of the organisms to be able to compare the results between studies, and to correctly generalise the findings to species level, and to understand the functional differences between the taxa in question.

Clitellata is a large “class” of segmented worms, comprising about one third of all known annelid species. It is placed within “subclass” Sedentaria (e.g., [17,18]), which is often thought of as a (major) polychaete group. Clitellates seem to have evolved in the transitional zones between marine and continental waters [19], and a majority of the species live within soil or aquatic sediment [20]. Unlike polychaetous annelids, they lack parapodia, and their prostomium lacks appendages. The monophyly of Clitellata is strongly supported by their unique mode of reproduction. Clitellates are hermaphrodites and characterized by the “clitellum”, an epidermal structure, secreting a protective cocoon for the embryos, which develop without a larval stage (see, e.g., [21]). The external morphology of clitellates is rather stable and offers few characters trustworthy of the taxonomic separation of taxa. The shape, position and number of gonads have historically been of fundamental importance for the classification [22]. The burrowing and interstitial habitats of most clitellates are likely to be the reason for their conserved morphology, as the evolutionary pressures in these environments may favour morphological stasis [3,9,23].
lack of externally discernible characters, many clitellates are hard to delimit and identify without the aid of molecular markers, and their species diversity has, in many cases, been underestimated when based on morphology alone (many examples will be given below). This fact has led to the rise of molecular approaches to separate species, which we will explore in this review.

Species delimitation can be divided into two steps, species discovery and species validation [24]. In the first step, the researchers form hypotheses about the species boundaries, which are then tested in the second step. In the species discovery phase, typically a single data source, e.g., morphology or DNA-barcoding, is used. Testing these hypotheses in the species validation step are often based on additional data and more sophisticated analyses. In most studies, this division between species discovery and validation is not explicitly stated, but rather implied.

The definition of cryptic species varies between researchers. Some use a relaxed definition. They count all cases as cryptic where species fall within the morphological variation of the same nominal species, even when there are minor differences between them, e.g., [3]. Others use a stricter definition and distinguish between true cryptic and pseudo-cryptic species, where the first refer to species between which no morphological differences are observed, while the latter are species that do show some differences, but still are so similar that they would be classified as the same nominal species based on morphology (e.g., [25]). In this paper, we apply the broader definition of cryptic species. Moreover, we use a liberal definition of molecular species delimitation. We include papers that explore molecular data to support species also discriminated morphologically, even if the authors do not explicitly test species limits.

In this paper, we aim to give an overview of the research field of species delimitation, and cryptic species, in Clitellata. We will examine the development of methods and the new data used in delimitation of clitellate species and discuss some of the problems arising when describing cryptic species. Finally, we will consider possible directions for this field.

2. History of the Field

Here we present an historical overview over the field of molecular species delimitation of clitellate worms, from a first publication in the 1970s to papers published in 2020. In total, 104 studies where found (Figure 1, Table S1). We identified four categories of data studied and have structured the overview accordingly, dividing this section into methods categorized as: (1) gel electrophoresis of proteins; (2) non-sequenced DNA; (3) Sanger-sequencing of a limited number of DNA fragments; (4) High-Throughput Sequencing (HTS) of a large number of DNA fragments. This classification is somewhat arbitrary, and methods from more than one category have been used together in many instances and is schematically shown in Figure 1.

![Figure 1. Historical timeline of the development of molecular species delimitation in Clitellata, showing the year of the first study, and the total number of studies, of the four major categories of methods referenced in this paper (see Table S1 for details). The histogram shows the total number of studies (all categories) per year.](image-url)
2.1. Protein Gel Electrophoresis

The first publications on the species delimitation of clitellates, by means of molecular data, explored variation in proteins revealed by gel electrophoresis. In these molecular methods, proteins encoded by alleles at some locus (alloenzymes), or proteins with the same function but encoded by separate genes at different loci (isoenzymes) are separated on gels, and the pattern observed is used to infer the separation of populations. The first works by using protein gel electrophoresis to explicitly test species hypotheses of clitellates that occurred in the 1970s and 1980s, (e.g., [26–28]), although the implication of this method was discussed already by Milbrink and Nyman [29], who saw it mainly as a supplement to morphological identification of species in ecological studies. Isoenzymes and alloenzymes continued to be used, often in combination with other methods (e.g., [30–35]). Another gel-based method is the study of general protein patterns, where a mix of proteins extracted from a specimen is run on a gel, producing a banding pattern that is then compared between individuals. The pattern produced is assumed to be species specific and an index based on protein patterns was suggested [36], which was then mainly used for studies of the family Enchytraeidae [30,33,35,37,38]. Crossed immunoelectrophoresis (CIE) is another method that, to our knowledge, was only tested once in clitellate systematics—i.e., to separate populations of Enchytraeus (Enchytraeidae) [39]. In general, these methods seem to have worked well, as the re-examination of the same groups using more modern methods has given similar results.

2.2. Non-Sequencing DNA Methods

Restriction Fragment Patterns [40] was an early DNA-based method for the separation of species, where restriction enzymes are used to digest specific markers and the variation in restriction fragments is visualised on a gel. It was used to separate species in the genus Enchytraeus (Enchytraeidae) [41]. A number of other methods that generate data on the presence/absence of amplification or length variation in markers, to estimate genetic variation, both within and between species, have been used in clitellate studies. These include Arbitrary Primers PCR (AP-PCR) [42], which uses a set of primers to amplify arbitrary genetic markers, and the presence or absence of amplification is scored and used as a measure of genetic distance. This method was applied by Koperski et al. [43] in a study on the leech Erpobdella octoculata (Erpobdellidae). The Random Amplification of Polymorphic DNA (RAPD) method [44] also amplifies random segments of DNA, but with several shorter primers. The amplified patterns are visualised on a gel and scored. This method was used in some studies [45–49]. In the Amplified Fragment Length Polymorphism (AFLP) method [50], DNA is digested by restriction enzymes, followed by the amplification of the fragments, which are then separated and visualised on a gel, and scored as absent/present. AFLP has been used in some papers [51–53]. Lastly, microsatellites [54–57] are short repetitive regions of DNA with a high mutation rate, and the variation within them can be studied both with and without sequencing. Microsatellites have been used to study gene flow between possible cryptic species in a few studies on lumbricid earthworms [58,59].

2.3. Sanger Sequencing

When proper DNA sequencing, i.e., the Sanger-sequencing method [60,61], became more affordable, it started to be used for the species delimitation of clitellates. The first studies (e.g., [62–64]) used a single mitochondrial marker and tried to find clusters of sequences divided by large genetic distances. Studies using a single marker are continuously published [65–81]. These studies still have their merits, especially when the analysis of single gene data is integrated with the examination of morphology or other independent information. Most of the single marker studies have either (1) been distance-based, identifying clusters of sequences with short genetic distances within each cluster, but greater distances between clusters, the so-called “barcoding gap”, i.e., a distinct gap in the distribution of genetic distances between low, i.e., intraspecific, distances and higher, i.e.,
interspecific, distances (see [82]), or (2) they have been tree-based, where a phylogeny is estimated, and used to identify well separated (monophyletic) clades, which are then being interpreted as potential species.

Today, however, studies based on more than one locus are becoming more and more common. In some analyses using multiple markers (e.g., [16,83–92]), the different sequence alignments are concatenated and a tree is estimated, and terminal clades are then identified and interpreted as species. Another approach is to estimate separate gene trees, or haplotype networks, and then identify congruent clades or network groups. Terminal clades (or specimen groups) found in all trees (or networks) are then interpreted as species, whereas conflicts between trees and groups are taken as support for gene-flow, and thus speak against speciation [13,93–115]. Several studies use a combination of the two approaches.

There is a plethora of software for dividing the individuals into species, as well as for testing species hypotheses. The most commonly used automated methods to divide single-marker datasets into species are Automated Barcode Gap Discovery, (ABGD) [116], and General Mixed Yule Coalescent (GMYC) [117]. ABGD delimits genetic clusters by detecting a significant gap in the pairwise distance distribution, and it uses genetic distances as the input. The method has been used in several studies [78,80,95,113,118–126]. GMYC, on the other hand, identifies a transition between the speciation and coalescence processes, by the identification of a shift in the branching patterns; the principle is that there are several short branches within species, but fewer and longer branches between species. It uses an ultrametric tree as input, i.e., a rooted tree where all terminal taxa are equidistant from the root; there is also a Bayesian implementation of the method (bGMYC), which applies Bayesian methodology, to account for uncertainty by sampling multiple trees [127]. This method has also been used for delimiting species of clitellates [78,113,118,126,128–131]. Another method is Bayesian Poisson Tree Processes (bPTP) [132]. It identifies significant changes in the pace of branching events on an input tree, using the number of substitutions between branching events, and it has been used in a few studies [118,131,133]. There are also a set of analyses in the Barcode of Life Database System (BOLD) [134], i.e., Barcode Gap Analysis (BGA) and the Refined Single Linkage (RESL) algorithm, the latter of which is the base of the Barcode Index Number (BIN) system [135]. These analyses have been used by Tiwari et al. [80] and Jeratthitikul et al. [118]. Haplowebs is a method that builds on the fields for recombination, i.e., sets of haplotypes connected by heterozygous individuals [136], where haplotype networks are constructed, and haplotypes that are found within the same heterozygous individual are connected to each other [137]. This method has been applied by Martinsson et al. [122] and Martin et al. [126].

To more formally test species hypotheses, both single and multi-locus approaches have been developed. Some of the single-locus methods are the statistical tests Rosenberg’s \(P_{AB}\) [138] and \(P_{(Randomly\ Distinct)}\) [139], which both test the distinctness of clades, and are implemented as a plugin in the software Geneious [140]. These tests have been used by some authors [119,121,123,124,131]. All of the methods mentioned in the previous two paragraphs are used on a single marker, and results from several loci have to be kept separate and each result interpreted as independent evidence. There are also explicit multi-locus species delimitation methods, and the most commonly used are based on the multispecies coalescent (MSC) model. In this model, genes evolve inside a species phylogeny where the branches are species and the properties of the branches restrict the gene trees. One of these restrictions is that the divergence times between species have to be more recent than the coalescent times for any genes shared between them, assuming no genetic transfer after speciation [141], and it can be used for the statistical testing of species assignments [2,142]. Different applications of MSC have been used in clitellate research, the most popular being the software BPP [143,144] used in several studies [12,113,122–125,145–147]. DISSECT (Division of Individuals into Species using Sequences and Epsilon-Collapsed Trees) [148], which is run within the software BEAST [149], is another species delimitation analysis based on the MSC and was used by Klinth et al. [119].
2.4. High-Throughput Sequencing (HTS)

An array of sequencing methods with a much higher throughput than Sanger sequencing have been developed today, and these methods are collectively known as Next-generation sequencing (NGS) or High-Throughput Sequencing (HTS). The techniques involved make the generation of genomic data possible, even for large samples of specimens, and HTS has made its way into species delimitation studies, also of clitellate worms. So far, four different methods have been used: (1) Restriction-Site-Associated DNA Sequencing (RAD-seq) [150,151] and (2) Genotyping by Sequencing (GBS) [152]; both work by using restriction enzymes for the digestion of the DNA, followed by the sequencing of short fragments from the restriction sites. This produces a dataset of DNA fragments from across the genome, which can either be used directly, or a set of Single Nucleotide Polymorphisms (SNP) and be extracted from the data and used for downstream analyses. The two methods differ mainly in RAD-seq implementing a fragment size selection step and more enzymatic and purification steps than GBS [152]. There are several variants of RAD-seq, and the double digest RAD-seq (ddRAD-seq) [153], which differs from the standard RAD-seq in that it lacks the random shearing and end repair of genomic DNA, but instead uses a double restriction enzyme digest, which reduces the cost of the library preparation, was used by Giska et al. [154]. On the other hand, Anderson et al. [155] use the standard RAD-seq protocol. Both of these studies are on the Lumbricus rubellus complex (Lumbricidae). GBS was used by Marchán et al. [156], to study the genus Carpetania (Hormogastridae).

(3) In Transcriptome Sequencing, the transcribed mRNA is being sequenced, and this generates a dataset consisting of expressed protein coding genes, which are then used for further analyses. Transcriptomes were used by Shekhovtsov et al. [157] and Shekhovtsov et al. [158] to study the Eisenia nordenskiold complex (Lumbricidae), but also in some larger phylogenomic studies [19,159,160]. (4) Anchored Hybrid Enrichment (AHE) [161] enriches the target region by using a probe for conserved anchor regions. This captures both the highly conserved anchor regions and the more variable flanking regions and enriches them in the sample before sequencing. AHE was used by Taheri et al. [147] to study Pontoscolex corethrurus (Rhinodrilidae), and Phillips et al. [162] to test hypotheses of leech evolution. The Whole genome sequencing of clitellates is still rare, and sequenced genomes only exists for a couple of species [163–165], and no phylogenomic studies focusing on Clitellata have used whole genomes.

3. Taxonomical Treatment of Delimited Species

As many nominal species have been found to actually be species complexes, each consisting of more than one species, the question arises, how should these species be treated taxonomically? Our opinion is that the species should as far as it is possible be described as such, and given a binominal name in the context of the traditional Linnean nomenclature. In many cases, delimited species have been described, either in the paper delimited them (e.g., [12,87,93,107,108,121,166,167]), or in subsequent papers with or without additional analyses [168–173]. However, we understand that this is not always possible, due to limited material, and nomenclatorial issues, etc. that prevent a description at the moment. One obstacle to overcome when revising a cryptic species complex is to determine which of the species should keep the original name, i.e., which species is identical with the type material used in the original description. This also needs to be done for any synonyms, as these names may be applied to other species in the complex. This work may be hard but is important for taxonomic stability. In cases where type material is missing, a neotype can be designated, and this has been done for some species (e.g., [12,71,81,113,169,174]). The problem with how to treat cryptic species has been discussed for Enchytraeidae [175], and the recommendations in that paper are largely valid across Clitellata (as well as for many other organismal groups) and are briefly summarised here. The main point is that description of new species should include a good morphological description, following the standard within the specific taxonomic group, if possible, combined with at least two genetic markers that are informative at the species level—e.g., 16S, COI, H3, or ITS—.
and at least one type specimen, preferably the holotype, should be sequenced. Further, specimens that are the basis for re-descriptions, including neotypes when appropriate, for nomenclatorial stability, should also be sequenced.

If species are delimited by genetic data in a study, and regardless of whether they are formally resolved, taxonomically or not, it is important that vouchers of the specimens used are deposited in natural history museums. This will enable the morphological re-examination of the specimens, to resolve possible conflicts between different datasets, as well as formal taxonomic description and revision.

4. Future Development of the Field

As we have shown in this overview, there is a great variation in the molecular methods used for species delimitation of clitellate worms, and we predict that the field will continue to grow and develop in the future. The recent introduction of High-Throughput Sequencing (HTS) methods in the systematics of clitellates has opened up a promising perspective, and we believe this will be commonplace in the near future. With continued methodological developments, we do not see a standardisation of methods used any time soon. However, there is a suggestion of using a standardised set of single-copy nuclear protein coding genes for species delimitation [176], which is an interesting suggestion, and perhaps, this will be developed and used in the future. It has the benefit of it being easier to re-use and combine data from more studies. We also see the great potential of Genotyping by Sequencing (GBS) as a relatively cheap method to generate genomic datasets for species delimitations—this method has already been used successfully for a group of hormogastrid earthworms [156]—and more studies using it will surely follow in the coming years. Finally, we hope that more of the delimited species will be formally described.

5. Summary and Conclusions

We hope this review has given a fair and inclusive description of how clitellate species have been delimited in recent years, thanks to a wide range of new data sources and methods, and also how we think delimited species should be handled and described from now on. Molecular species delimitation of clitellate worms is a research field in constant movement, evolving with molecular systematics at large that of course is universal to all groups of organisms, and we see no signs for this development to slow down. We hope this paper will give inspiration to further studies and the exploration of new methods.

With the continued testing of the many species hypotheses in Clitellata, characterized by a population genetics approach rather than traditional analyses of similarities and differences, we will get a better understanding of the species taxonomy of this species-rich and common annelid group. This will improve other fields of clitellate biology, especially with regard to phylogeny (evolutionary history) and classification, and it may stimulate studies on more applied aspects of their biology and function in various ecosystems (as suggested by [9]).

Supplementary Materials: The following are available online at https://www.mdpi.com/1424-2818/13/2/36/s1, Table S1: List of studies using molecular data to delimit species of Clitellata, listed chronologically.

Author Contributions: Conceptualization, S.M. and C.E.; writing—original draft preparation, S.M.; writing—review and editing, S.M. and C.E.; funding acquisition, C.E. All authors have read and agreed to the published version of the manuscript.

Funding: The study was funded by the Swedish Taxonomy Initiative, Swedish Species Information Centre (SLU ArtDatabanken); the Norwegian Taxonomy Initiative, Norwegian Biodiversity Information Centre (NTNU Artsdatabanken); the Swedish EPA’s Environmental Research Fund in collaboration with the Swedish Agency for Marine and Water Management.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

4. Pfenninger, M.; Schwenk, K. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evol. Biol. 2007, 7, 121. [CrossRef]
24. Carstens, B.C.; Pelletier, T.A.; Reid, N.M.; Satler, J.D. How to fail at species delimitation. Mol. Ecol. 2013, 22, 4369–4383. [CrossRef]
Diversity 2021, 13, 36


85. Shekhovtsov, S.V.; Golovanova, E.V.; Peltek, S.E. Genetic diversity of the earthworm *Octolasion tyrtaeum* (Lumbricidae, Annelida). *Pedobiologia* 2014, 57, 245–250. [CrossRef]

86. Shekhovtsov, S.V.; Rapoport, I.B.; Poluboyarova, T.V.; Geraskina, A.P.; Golovanova, E.V.; Peltek, S.E. Morphotypes and genetic diversity of *Dendrobaena schmidtii* (Lumbricidae, Annelida). *Vavilov J. Genet. Breed.* 2020, 24, 48–54. [CrossRef]


115. Felföldi, T.; Dózsa-Farkas, K.; Nagy, H.; Hong, Y. Three new enchytraeid species (Enchytraeidae, Annelida) from mountain soils of Korea and ten species new for the country. *Zootaxa* 2020, 4896, 1–45. [CrossRef]


120. Latif, R.; Malek, M.; Csuzdi, C. When morphology and DNA are discordant: Integrated taxonomic studies on the *Eisenia fetida/anatroecis* complex from different parts of Iran (Annelida, Clitellata: Megadrili). *Eur. J. Soil. Biol.* 2017, 81, 55–63. [CrossRef]


127. Reid, N.M.; Carstens, B.C. Phylogenetic estimation error can decrease the accuracy of species delimitation: A Bayesian implementation of the general mixed Yule-coalescent model. *BMC Evol. Biol.* 2012, 12, 196. [CrossRef] [PubMed]


