DNA taxonomy and barcoding of monogenean parasites: lessons from Gyrodactylus

Haakon Hansen, Tor A. Bakke and Lutz Bachmann

Natural History Museum, Department for Zoology, University of Oslo, PO Box 1172, Blindern, NO-0318 Oslo, Norway

DNA taxonomy and barcoding use nucleotide sequence data to achieve comprehensive species descriptions that facilitate reliable species diagnostics and rapid assessment of biodiversity, both of which are of great importance for parasitologists. Such molecular approaches have been applied to the monogenean genus Gyrodactylus, in particular to G. salaris, the cause of serious gyrodactylosis on Atlantic salmon. Here, we discuss, using the example of G. salaris and related species, why DNA barcodes, although powerful for biodiversity assessment, are insufficient to appropriately characterize parasite species – from a parasitological point of view – in the absence of additional data on and infection biology and morphology.

Gyrodactylus diversity and DNA taxonomy and barcoding

The monogenean flatworm genus Gyrodactylus currently includes ~400 morphologically highly conserved species that are ectoparasitic on marine and freshwater fish [1]. Based on the assumptions of relatively strict host specificity and at least one Gyrodactylus species infecting each fish species, estimates of as many as ~20,000 Gyrodactylus species have been put forward [2]. However, the current rate of Gyrodactylus species discovery of only 20–30 new species per decade [3] is very low. The concepts of DNA taxonomy and DNA barcoding (see Glossary) can facilitate unambiguous species identification, diagnoses of problematic taxa and discovery of ‘cryptic’ species [4–10]. The main arguments in favour of DNA taxonomy and barcodes, – high conservation of morphology, which makes it difficult to distinguish related species, and the low throughput of classical taxonomy – apply perfectly for Gyrodactylus. The prospect of a rapid assessment of parasite biodiversity and better tools for diagnostics for these morphologically very similar species seems very promising, because unambiguous identifications of the causes of parasitoses are essential for proper disease management.

Sequencing of the mitochondrial cytochrome oxidase I (COI) gene, the recommended marker for barcoding animal species, revealed that the biodiversity of Gyrodactylus is considerably underestimated [11–14]. However, the example of the notorious Atlantic salmon (Salmo salar) parasite G. salaris Malmberg, 1957, and the closely related G. thymalli Žitnáň, 1960, found on grayling (Thymallus thymallus), illustrates the need for further biological characterization of the parasites themselves in addition to the molecular barcodes. Assessment of biodiversity through molecular tools is certainly welcome, but COI barcodes alone cannot support meaningful management because they do not provide information about important epidemiological parameters such as, for example, infectivity and virulence of species or strains.

Gyrodactylus salaris and gyrodactylosis

Parasitologists must identify, classify and discriminate between virulent and non-virulent pathogens unambiguously. Tools that are diagnostically and biologically meaningful are welcomed by those responsible for disease control. Details on the biology of most Gyrodactylus flatworms are, to a large extent, unknown. These progenetic parasites are hermaphroditic, partly parthenogenetic and hyperviviparous [3]. Thus, a single parasite could establish a new metapopulation. Only few Gyrodactylus–host associations, such as G. salaris on Atlantic salmon and rainbow trout (Oncorhynchus mykiss) [2], have been studied in

Glossary

Hypervivipary: an exceptional progenetic reproduction mode that resembles ‘Russian dolls’: a mother contains in utero a gravid daughter with another developing embryo within. This mode of reproduction is derived from oviparity and is found in many Gyrodactylus species.

Well-supported clade: various methods of objective phylogenetic inference yield evolutionary trees for visualizing the phylogenetic relationships of taxa. A clade is a group of closely related taxa consisting of a single common ancestor and all the descendants of that ancestor. Thus, a clade is a scientific hypothesis of evolutionary relationships among taxa. Methods that resample a particular dataset, for example with replacement (the bootstrap), can provide confidence that particular groupings of clades are robust given the data available.

Mitochondrial haplotype: the genetic constitution of a genetic locus, a chromosome or a genome is referred to as a haplotype. In the absence of recombination, a whole mitochondrial genome represents a single haplotype. This makes mitochondrial sequences useful tools for studying the evolutionary history of populations.

DNA taxonomy: nucleotide sequences are, as a first approximation, taken as a tag for the identification of the species that was the source for the extracted genetic material. These sequences are deposited in public databases, and the DNA extracts serve as molecular vouchers to be stored in natural history collections.

DNA barcoding: the diversity of life can be discriminated through the analysis of a small segment of the genome, in animals a stretch of the mitochondrial cytochrome oxidase I (COI) gene. This promising approach to the diagnosis of biological diversity has gained broad acceptance in the scientific community. Currently, the major Barcoding of Life projects focuses on fish, birds and Lepidoptera.
Taxonomy and systematics of *Gyrodactylus salaris* and *G. thymalli*

The classical description of species of the genus *Gyrodactylus* is based on the morphology of the hard parts (sclerites) of the posterior attachment apparatus (opisthaptor) in addition to host preference (Box 1). Unfortunately, morphological discrimination of species such as *G. salaris* and the closely related non-pathogenic *G. thymalli*, if possible at all, requires sophisticated statistics [17,18]. Recently applied molecular markers, such as the ribosomal internal transcribed spacers (ITS1 and ITS2), the intergenic spacer (IGS) and the COI gene, have yielded somewhat contradictory results. The ITS regions, frequently used as markers in DNA taxonomy, are identical for most *G. salaris* and *G. thymalli* populations, which could be taken as an indication of conspecificity [19]. A first screening of the repeat arrangements within the two repetitive IGS regions identified a ‘*salaris* form’ and a ‘*thymalli* form’ but neither form is truly diagnostic for discriminating *G. salaris* and *G. thymalli* [13,20,21]. The best resolution to date has been observed with COI-based mitochondrial haplotypes [11–14].

**Box 1. Commonly used markers in *Gyrodactylus* taxonomy and systematics**

**Morphological markers**

Most descriptions, delimitation and discrimination of *Gyrodactylus* species have been based on morphology and host specificity. The hard parts (sclerites) of the posterior attachment organ (opisthaptor), which consists of the ventral bar, two hamuli and 16 marginal hooks, have been the most important traits for morphological analyses. Linear and angular distance measurements are applied for such purposes. For some species, such as *G. salaris* and *G. thymalli*, morphological discrimination requires sophisticated statistical analyses of the morphological data, and the results are of questionable relevance. Abiotic environmental factors, such as ambient temperature, also affect the size of the sclerites, which could render morphometric studies on *Gyrodactylus* species even more prone to error.

**Molecular markers**

In recent years, the internal transcribed spacers (ITS1 and ITS2) of the nuclear ribosomal gene cluster have been used for identification of *Gyrodactylus* species. As of December 2006, a GenBank search using the keywords ‘*Gyrodactylus*’ and ‘internal transcribed spacer’ together returned 188 hits for ~80 species. Most species have specific ITS sequences, and it has been suggested that a sequence diversity of 1% between species indicates species status. The repeat region of the intergenic spacer (IGS) of the nuclear ribosomal gene cluster has also been used as a molecular marker in studies on *G. salaris* and *G. thymalli*. Early studies on the repeat arrangements within the two repetitive IGS regions have led to the description of a ‘*salaris* form’ and a ‘*thymalli* form’, which have been considered informative for discriminating *G. salaris* and *G. thymalli*. However, more recent studies based on extended sampling indicate that neither form is fully diagnostic. Even repeat arrangements with one repetitive IGS region similar to the *salaris* form and the other repetitive IGS region similar to the *thymalli* form have been found within the same specimen. In the case of studies on *G. salaris* and *G. thymalli* it was shown that the mitochondrial cytochrome oxidase I (COI) gene can reveal considerably higher levels of diversity than nuclear ribosomal markers. These studies support the idea that COI-based barcoding is promising for assessing monogenean biodiversity, albeit with certain applied limitations.

COI diversity in *G. salaris* and *G. thymalli*

Comprehensive surveys of the genetic diversity of *G. salaris* and *G. thymalli* throughout Northern and Central Europe to find diagnostic markers have revealed several well-supported clades of mitochondrial haplotypes (haplogroups) [11–14]. In particular, the *G. thymalli* haplogroups reflect the geographic origin of the parasite strains, but there is no support for the monophyly of either *G. salaris* or *G. thymalli*. In recent years, an extensive dataset consisting of almost 250 available COI sequences, corresponding to 44 haplotypes of numerous populations from four different fish hosts, has been generated [11–14] (Figure 1). In the neighbor-joining dendrogram shown, three haplogroups refer to *G. salaris* and 12 haplogroups to *G. thymalli*. Some clades consist of only one haplotype. Most of the well-supported clades consist of parasites from either one host and/or one watercourse. There are, for example, two haplogroups that relate to *G. thymalli* of the Norwegian Glomma and Trysil watercourses, respectively. One *G. salaris* clade, however, which consists of only haplotype F, includes parasites from at least three host species: salmon, rainbow trout and Arctic charr (*Salvelinus alpinus*) [11,22]. Parasites with haplotype F also show different infectivity and population growth in infection experiments [2,23]. Thus, there is no strict congruence between mitochondrial sequence diversity and parasite virulence, host specificity and the current delimitation of species. What, then, is the utility of these barcodes for *G. salaris* and *G. thymalli*?

COI-based species and strain identification in *G. salaris* and *G. thymalli*

It has been suggested that ~650 bp of the 5′ end of the COI gene can serve as a universal barcode marker for animals [9,10]. In brief, DNA barcoding projects intend to sequence COI for as many species as possible and preferably all species of each higher order taxon under study to provide a framework for a rapid molecular identification [4]. It then follows that biodiversity can be assessed by defining molecular operational taxonomic units (MOTUs) [7,24] but without the need for comprehensive species descriptions. This strategy has been successfully applied for many problematic taxa [5,7,25–27]. Indeed, in *Gyrodactylus*, analyses of ~800 bp of the COI gene that largely cover the aforementioned universal barcoding marker have revealed more diversity than morphology or other commonly used molecular markers such as ribosomal ITS sequences [11,12]. Furthermore, the COI data are also informative for tracking the invasion history of *G. salaris* into wild Atlantic salmon and farmed rainbow trout populations in Fennoscandia and Russia [11,12]. The high number of well-supported clades of mitochondrial haplotypes might suggest considerably more *Gyrodactylus* species than estimates based on the calcu-
Figure 1. Phylogenetic relationships of mitochondrial haplotypes of *Gyrodactylus salaris* and *G. thymalli* based on sequences of the cytochrome oxidase I (COI) gene. Approximately 820 bp of the gene have been amplified using published primer pairs [11–14]; this fragment covers most of the 5' end of the gene, suggested by others [9,10] as the universal animal barcoding marker. The neighbor-joining analysis [36] is based on a trimmed global alignment of 745 bp of each haplotype to avoid missing information. The neighbor-joining dendrogram includes all currently known haplotypes of *G. salaris* and *G. thymalli* and is based on K2P-distances [28], which is the typical method used for DNA barcoding projects. Bootstrap support estimates (shown on selected branches were obtained by running 1000 resampling replicates. The haplotypes have been identified by compiling 238 published sequences of *G. salaris* and *G. thymalli* [11–14] and/or are available through GenBank. Altogether these sequences represent 44 haplotypes. Clades are defined as haplogroups that receive at least 80% bootstrap support in the neighbor-joining analysis. Haplotypes isolated from grayling, presumably *G. thymalli*, are in green and those from Atlantic salmon, presumably *G. salaris*, in red. Clades consisting of *G. thymalli* haplotypes are labelled according to geographical origin of the parasites or the respective drainage system. For the majority of *G. salaris* haplotypes (i.e. haplotypes A–F and those starting with ‘Sal’), an assignment to a particular drainage system is not possible. Haplotypes that group together in clades are usually found on only one host. However, haplotype F (black) has been detected in parasites isolated from Atlantic salmon, Arctic charr and rainbow trout. The scale bar indicates 0.025 substitutions. Country abbreviations: DK, Denmark; FIN, Finland; LAT, Latvia; NOR, Norway; POL, Poland; RUS, Russia; SLO, Slovenia; SWE, Sweden; UK, United Kingdom.

**Gyrodactylus lavareti**, AY225306

**Gyrodactylus sp.** AY258375

TRENDS in Parasitology
lations described earlier [2]. Following a rule of thumb used in other taxa, that COI sequence diversity exceeding $\sim 2\%$ is an indication of separate species status whereas sequence diversity $< 2\%$ reflects intraspecific variation [10], G. salaris and G. thymalli should certainly be split into several species. The average genetic two-parameter distance [28] between all haplogroups depicted in Figure 1 is 0.0222 with 0.0181 for comparisons including only G. salaris clades, 0.0225 for comparisons including only G. thymalli clades, and 0.0183 when comparing G. salaris with G. thymalli clades. In fact, a split of G. salaris and G. thymalli into several species according to the phylogenetic species concept has been suggested as one possible taxonomic conclusion from the observed COI diversity [11]. However, fixed threshold values are always problematic and can at best be considered guidelines. Furthermore, it has been shown that threshold values applied to species discovery can be error prone. For 236 evolutionary significant units (ESU) of cowries, 16% of the ESUs were artificially grouped at a 3% threshold for species differentiation, owing to the considerable abundance of young taxa [29]. By contrast, the 3% threshold for intraspecific variation was correct for 98% of taxa [29]. In fact, for G. salaris and G. thymalli it has also been suggested to either consider them one polytypic species, or to keep the status quo with both as two valid species discriminated primarily by different host preferences [2,20,30].

At first glance, this point could be considered an academic question. However, splitting species into several taxonomic units defined by considerably differentiated mitochondrial haplogroups will result in a taxonomic inflation that might be acceptable for systematicists but is certainly a ‘worst case scenario’ for pest management purposes. Clear and easily understood guidelines for identification of pathogens are then almost impossible to establish but nevertheless mandatory for efficient management of notifiable diseases. For the same reasons, lumping together harmless and virulent parasite forms, although favoured by some taxonomists [12], would also be very inconvenient (if not unacceptable) for pest management and international legislative purposes. However, a purely typological or biological species concept that is based on factors such as host preferences has substantial shortcomings as well. Although it might enable easy identification of the cause of a parasitosis, it will fall short when it comes to setting up countermeasures for controlling and/or eradicating the parasites. Potential sources of infection or transmission could simply be overlooked by such an a priori exclusion.

A more detailed look at G. salaris and G. thymalli molecular diversity further illustrates the problem. According to the neighbor-joining dendrogram depicted in Figure 1, 15 mitochondrial lineages or haplogroups can be identified. Neglecting the possibility that expanded sampling might lead to the collapse of some clades and new discovery of others, parasites belonging to only three of the clades are known as agents of severe gyrodactylosis. The parasites in these clades are all classified as G. salaris. However, the particular haplotype F has been detected not only on rainbow trout and as a severe pathogen on Atlantic salmon [2], but also on Arctic charr in a South Norwegian Lake, where no harm has been reported [22]. Later experimental laboratory studies have shown that the parasites from Arctic charr cannot severely infect Atlantic salmon [23]. Nevertheless, Norwegian authorities were informed once the parasites had been detected, and countermeasures and restrictions on fishing were immediately imposed. Concern was raised by local fishermen because they perceived, correctly as it is now known, that the parasites on Arctic charr were harmless. Nevertheless, the authorities were correct to implement restrictions because the harmless nature of the parasites remained unclear until time-consuming infection experiments had been conducted [23]. The occurrence of G. salaris on Arctic charr and not on Atlantic salmon was no reason for an ‘all clear’, because it has been shown that highly virulent G. salaris can reproduce on Arctic charr [31] and survival on wild Arctic charr is also observed in the North Norwegian rivers Signaldalselva and Skibotnelva [32,33].

As Figure 1 illustrates, well-supported clades or haplogroups can be identified on the basis of COI sequences. Unfortunately, the phylogenetic relationships between these haplogroups remain unclear because bootstrap resampling statistics indicate poor support for particular main tree arrangements. This makes it impossible to identify the closest relatives of the haplogroups that relate to G. salaris, that is, the parasite that causes the notifiable gyrodactylosis. If one assumes that the notorious G. salaris results from a host switch from a previously harmless parasite strain, it would be of utmost importance to identify the strains that might bear the highest risk for further host switches in the future. Host switches are assumed to be common in Gyrodactylus and to be a major reason for speciation [19]. However, the COI data do not allow us to recognize strains with potential for host switches. On the other hand, the COI data help to define clades of strains that can be experimentally tested for their host switch potential.

It should be emphasized that there is currently no other marker available with such a high resolution for characterizing strains. Attempts have been made to analyze morphological differences between the strains of G. salaris and G. thymalli, which carry significantly different mitochondrial haplotypes [18]. Unfortunately, such analyses are very difficult, as abiotic effects such as macroenvironmental differences in temperature are known to affect the morphology of the sclerites [34,35]. To collect and keep parasites to be compared under identical macro- and microenvironmental conditions in the laboratory is best very time-consuming, if even practical at all. Preliminary data (K. Olstad, personal communication) indicate that morphometric discrimination of haplogroups of G. salaris and G. thymalli sampled in the wild is impossible. As such, sequencing of COI is currently the fastest way to diagnose specific G. salaris and G. thymalli populations.

Conclusions

Without doubt, the example of G. salaris and G. thymalli illustrates that COI sequences can detect substantial genetic diversity of monogenean ectoparasites. DNA barcoding approaches can therefore be recommended to assess the biodiversity of largely unknown and problematic taxa,
such as *Gyrodactylus*. To some extent, COI barcodes also provide important information for taxonomy, systematics, biogeography and ecology. However, for a meaningful delimitation of relevant taxonomic units that facilitate the management of parasite induced diseases, a purely COI-based molecular classification system is not satisfactory. The crux of the matter is that *G. salaris* and *G. thymalli* COI barcodes do not correlate diagnostically with host preference and virulence of strains. As such, from the point of view of pest management, whether or not such taxonomic units relate to a particular species concept and/or a particular taxonomic rank seems to be of minor importance.

References

15. McHugh, E.S. et al. (2000) Discrimination of the notifiable pathogen *Gyrodactylus salaris* from *G. thymalli* (Monogenea) using statistical classifiers applied to morphometric data. *Parasitology* 121, 315–323

www.sciencedirect.com