# GYRODACTYLIDS ON EUROPEAN SALMONID HOSTS: GENOMICS, PHYLOGEOGRAPHY AND SPECIATION

Christoph Hahn


Dissertation presented for the degree of

Philosophiae Doctor (PhD)

Natural History Museum

University of Oslo

2014

**Supervisor**

Lutz Bachmann
Natural History Museum
University of Oslo
Norway

**Co-supervisor**

Tor A. Bakke
Natural History Museum
University of Oslo
Norway

**Co-supervisor**

Phil D. Harris
Natural History Museum
University of Oslo
Norway

**External supervisor**

Steven Weiss
Institute of Zoology
Karl-Franzens-University Graz
Austria

**Adjudicating committee**

***1st opponent***

Tine Huyse
Department of Biology
Royal Museum for Central Africa
Belgium

***2nd opponent***

Walter Salzburger
Institute of Zoology
University Basel
Switzerland

***Administrative leader***

Emily Rebecca Cramer
Natural History Museum
University of Oslo
Norway

# CONTENTS

# PREFACE

"The drama's done. Why then here does any one step forth? Because one did survive the wreck."

- Herman Melville, *Moby-Dick*

The document you hold in hand (or maybe view on your screen) marks the conclusion of an important phase of my life. I remember feeling fascinated, frustrated, exhausted, excited, hungry, surprised, .. - it was a tremendous journey indeed. Thinking back, I like to believe that I have grown, both as a person and as a scientist. I learned a lot about many things - about Nature and Science, about integrity, and about people, including myself. Over the years many people have contributed to this work and to my development. Some came and went - others were with me all along. Some contributed directly - some indirectly. It is both likely and unfortunate that I have not always expressed my gratitude for the smaller or larger part that you have played in this journey of mine. Whatever it was, I wouldn't be who and where I am today without it and finally, this is the time and place to thank you all – in writing!

I'd like to begin with my supervisors: Lutz Bachmann, you were an incredible resource of knowledge - molecular biology far from being your only area of expertise. I am certain that your enthusiasm had a large effect on the success of my work and I really appreciate you giving me the slack to venture into the unknown (and for giving me a push sometimes when I got myself lost in details). I remember as if it was yesterday, when Tor Bakke sat down with me for the first time to introduce me to *Gyrodactylus* – since then it has "grown on me" to an extent that I wouldn't have believed – thanks for "infecting" me. Phil Harris was a constant resource of ideas. You always had a lot of time for my thoughts, opinions and doubts regarding my scientific- or personal development. I appreciate all the lively discussions we had, they were crucial for my understanding. Steven Weiss has been a constant in my scientific life since my bachelor studies – through long field trips and many fruitful discussions - I consider myself lucky for that.

Work at the museum wouldn't be the same without the familiar faces one sees every day. Thanks to Ann-Helen, Ania, Bastian, Eve, Joost, Paula and Susann for the cakes, lunches,

coffees and chocolate we shared and the generally pleasant working environment you helped creating.

A few people even managed to become a big part of my personal life – my Oslo experience would not have been the same without you. Raul, Audun, Ingunn, Judith and Manuel – thanks for the good times, the beer brewing, barbecuing, fishing, concerts, hytte trips, dinners, etc. I am proud to call you my friends! Looking forward to the visits!

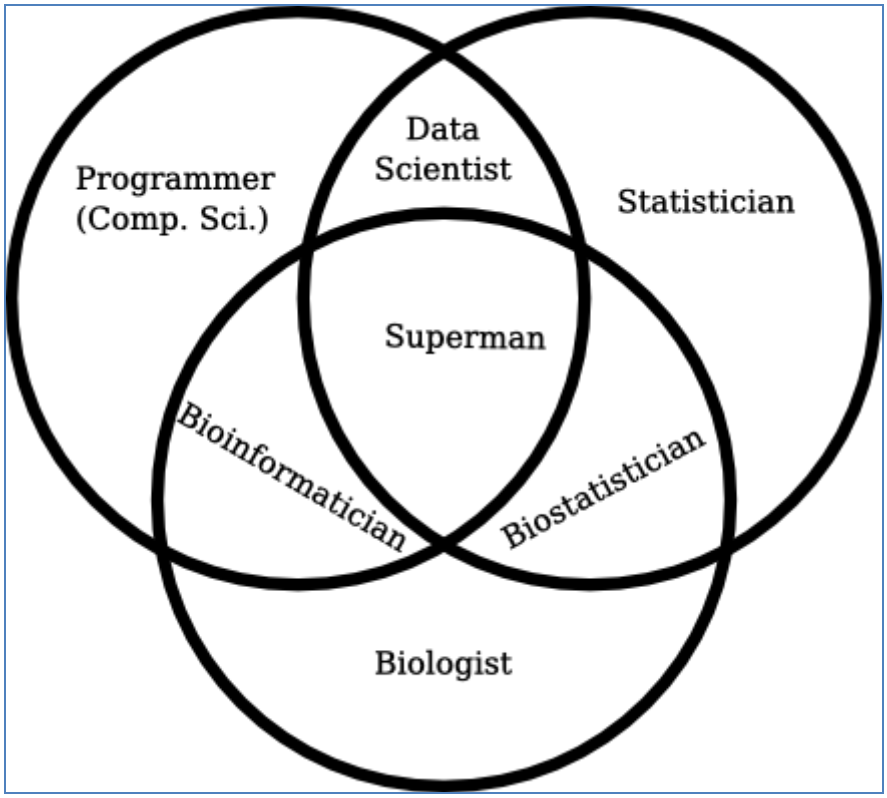
Thanks to Bastien for letting me pick your brain about MIRA and the subtleties of NGS data.

I would like to thank a number of people who I was lucky enough to do field work with over the years – some even taught me the secrets of electro-fishing (in alphabetical order): Dario, Edgar, Etienne, Gerhard, Gü, Günter, Harri, Henri, Javier, Josef, Kurt, Peter, Stojmir, Thomas, Viktor, Wolfi and Wolfi. I enjoyed every single day and I sincerely hope there will be opportunities in the future!

I am grateful that my grandparents supported my decision to try and do something that I liked instead of something that was likely going to get me a job. I wish you could see me now! The occasional visit to Austria wouldn't have been so superb without the constant support, the great food and drink, and general hospitality of our families – you are amazing! I am glad that despite the distance we managed also to keep in contact with some of our oldest friends – you have been another highly appreciated constant - always a pleasure to see you - in Austria, Norway, or wherever..

My precious wife Agnes, I will be ever grateful that you have joined me in this Norway adventure. At times it was dark and cold, but together we endured. At times it was lonely, because we were far apart, yet you were always with me in my thoughts, and together we endured. That all sounds very dramatic, but the bottom line is that we made it, together, through exciting times full of ups and downs and full of change; the most important being our sweet Olivia. I could have never asked for more!

Christoph Hahn, 2014




Fejes, A.P. 01/2014 (http://blog.fejes.ca/?p=2418)

## LIST OF PAPERS

The current thesis is based on the following four papers, which will be subsequently referred to in the text by their respective roman numerals.

I. **Hahn C.**, Bakke T.A., Bachmann L., Weiss S., Harris P. (2011) Morphometric and molecular characterization of *Gyrodactylus teuchis* Lautraite, Blanc, Thiery, Daniel & Vigneulle, 1999 (Monogenea: Gyrodactylidae) from an Austrian brown trout population. *Parasitology International*. 60(4):480-7. doi: 10.1016/j.parint.2011.08.016.

II. **Hahn C.**, Bachmann L, Chevreux B. (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research.* 41(13):e129. doi: 10.1093/nar/gkt371.

III. **Hahn C.**, Fromm B., Bachmann L. (2014) Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparastic Neodermata. *Genome Biology and Evolution*. 6(5):1105-17. doi: 10.1093/gbe/evu078.

IV. **Hahn C.**, Weiss S., Bachmann L. Co-speciation of the ectoparasitic *Gyrodactylus teuchis* (Monogenea, Platyhelminthes) and its salmonids hosts. Manuscript.

# ABSTRACT

The present PhD thesis, entitled "Gyrodactylids on European salmonids hosts: genomics, phylogeography and speciation", presents a number of significant contributions to the field of evolutionary biology. It offers novel insights into mechanisms of evolution (**PAPER III**, **PAPER IV**), introduces methodological advances (**PAPER II**) and, overall, provides important foundations for future research (**PAPER I**, **PAPER II**, **PAPER III**, **PAPER IV**). My work has focused on the biology and evolution of ectoparasites of the genus *Gyrodactylus* infecting European salmonids and I have applied a variety of methods including large scale bioinformatics approaches. *G. salaris*, along with its benign sister species *G. thymalli*, has been in the center of gyrodactylid research for over 30 years, since the first outbreak of the Norwegian salmon epidemics. *G. teuchis* and *G. truttae* are non-pathogenic, widely distributed, and largely neglected parasites, which bear striking morphological similarity and overlap in host range with the enigmatic *G. salaris.* My initial paper demonstrates that the so-called cryptic species *G. teuchis* can indeed be discriminated from the latter based on a Principal Component Analysis (PCA) of 32 morphometric characters. The *G. teuchis* population in focus represents the first natural record from a Danubian trout population. We furthermore describe genetic variation in the ribosomal internal transcribed regions, which we interpret as evidence for a non-fully homogenized ribosomal DNA cluster, a possible result of recent introgression. Furthermore, I have devised a novel *in silico* approach for the reconstruction of animal mitochondrial genomes directly from genomic Next Generation Sequencing (NGS) data. The method was developed with particular emphasis on gyrodactylids and its applicability is demonstrated for two *Gyrodactylus* species, and their respective hosts, based on real and simulated Illumina data. A further paper presents the first high quality draft genome for *G. salaris*, representing an important platform for future research. We resolve the disputed interrelationships of the three major parasitic flatworm groups with a large scale phylogenomic approach and find the ectoparasitic Monogenea basal to the endoparasitic tapeworms and flukes, indicating ectoparasitism as the ancestral state in the obligate parasitic Neodermata. Comparative genomic analyses of seven parasitic flatworm genomes identify a number of shared genomic features between endo- and ectoparasitic lineages. In a final paper, I present firm evidence for the co-speciation of *G. teuchis* with its salmond hosts brown trout and Atlantic salmon, based on extensive European sampling and

co-phylogenetic analyses. Although co-speciation represents an intriguing concept in evolutionary biology it has rarely been demonstrated in the literature.

# 1. INTRODUCTION

## 1.1 GENERAL INTRODUCTION TO *GYRODACTYLUS*

The hyperdiverse genus *Gyrodactylus* v. Nordmann, 1832 (Platyhelminthes, Monogenea) currently includes about 400 parasitic species (Harris et al. 2004) on teleost fish, but this number is thought to poorly reflect the diversity of the radiation, estimated to be at least ~ 20,000 species (Bakke et al. 2002). Most species infect the skin and fins of their hosts, but some also occur on the gills. Generally gyrodactylids are epidermal browsers that occasionally take dermal cells as well (Bakke et al. 2007). The direct gyrodactylid life cycle is characterized by hyperviviparity, a rare but highly efficient reproductive mode (Cohen 1977). Individual worms retain fully grown daughters *in utero* until they themselves contain developing embryos (see Figure 1). This first embryo always develops asexually (Cable and Harris 2002). Gyrodactylids are highly progenetic (Harris 1983). Progenesis is defined as the acceleration of the life cycle to allow an organism to reproduce at a larval or juvenile stage (Bakke et al. 2007). Newly born daughters of *Gyrodactylus* can give birth after only days (Jansen and Bakke 1991), which potentially enables rapid population growth (Cable and Harris 2002). Gyrodacytlids are protogynous hermaphrodites; the male reproductive system only develops after the first daughter is born. Second and subsequent daughters may originate from sexual or asexual reproduction. The relative proportion of sexual and asexual offspring in natural populations of *Gyrodactylus* is likely to be species specific and may furthermore depend on the host species or even the particular host individual. Mating was observed in *G. turnbulli* (Harris 1989; Schelkle et al. 2012) and sexual reproduction was speculated to be important during phases of epidemic population growth (Harris 1989), while experimental evidence suggests asexual reproduction to be dominant in *G. gasterostei* (Harris 1998). Using an experimental approach Harris et al. (1994) inferred regular sexual reproduction in *G. salaris*. In contrast, purely asexual reproduction has been postulated for a putative triploid lineage of *G. salaris* (Zietara et al. 2006), but karyological and/or unambiguous molecular evidence for this claim has yet to be presented. Water temperature has been shown to have a marked effect on reproductive rate and survival of *Gyrodactylus* (Jansen and Bakke 1991, 1993; Olstad et al. 2006), and might thus indirectly affect population structure. A significant increase in the relative proportion of asexual first births was observed as a result of decreased survival time caused by sub-optimal water temperature in *G. gasterostei* (Harris 1993).

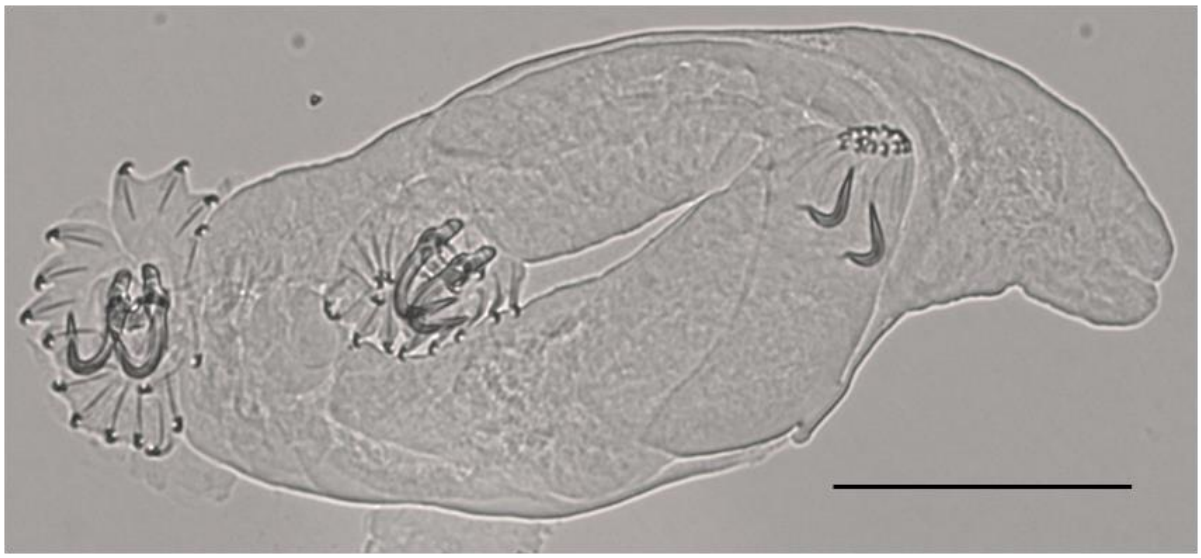


**Figure 1**: Microscopic image of *G. teuchis* collected from *S. trutta*, displaying hyperviviparity. Scale bar = 100 µm.

*Gyrodactylus* transmission occurs directly from host to host, without any intermediate stages. The direct gyrodactylid lifecycle is usually completed on one single host individual and detachment from the host is usually fatal within 24h hours (reviewed in Bakke et al. (2007)). Gyrodactylids attach to their hosts mainly via the opisthaptor, a posterior structure containing hard parts. The morphology of these parts, i.e. marginal hooks, hamuli and bars (see Figure 2), along with information about the host species were traditionally used to characterize *Gyrodactylus* species. Not surprisingly, purely host based *Gyrodactylus* taxonomy insufficiently reflects the real diversity in the group and can be misleading as distinct species often overlap in their host range (e.g. *G. teuchis* and *G. truttae*; see section 1.2.2 below). Morphological species identification is difficult due to the relative morphological conservatism and paucity of reliable landmarks within the group. Morphometric approaches have proved relatively robust (Shinn et al. 2004), but recently Vignon (2011) argued for the need to move on from traditional morphometrics and focus on the shape of morphological features in the classification of gyrodactylids instead (see also Olstad et al. (2009) for an examination of shape differences in *Gyrodactylus*). However, the supremacy of molecular species identification over morphological approaches for gyrodactylids has been recently demonstrated (Shinn et al. 2010). Whatever the method of choice, accurate species identification relies on the availability of reliable reference data, including sequences, and is not just an academic exercise, as some species are notifiable pathogens that can cause devastating ecological and economic damage (e.g. *G. salaris*, see section 1.2.1 below).

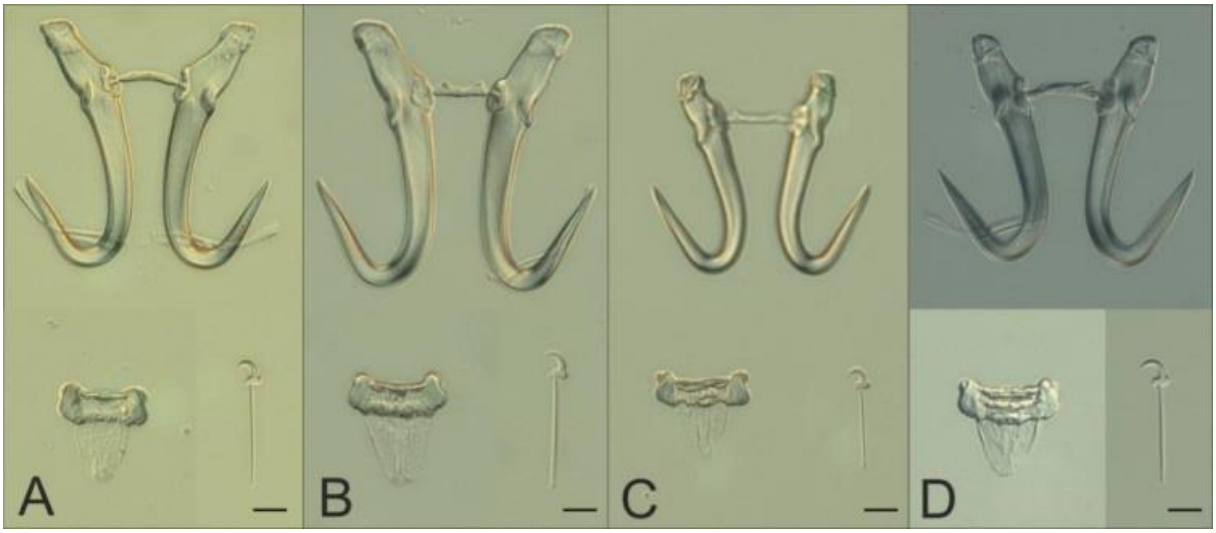



**Figure 2:** Hamuli (top), ventral bars (bottom left) and marginal hooks (bottom right) of four *Gyrodactylus* species. (A) - *G. salaris*, (B) - *G. thymalli*, (C) - *G. truttae*, (D) - *G. teuchis*. Scalebar = 10µm.

Modern species descriptions usually contain both morphological and molecular data (see e.g. King et al. (2014)). The principal molecular markers in *Gyrodactylus* taxonomy are currently the ribosomal ITS regions (Internal Transcribed Spacers) and the mitochondrial Cytochrome Oxidase I (COI) gene. These markers in many cases provide robust and unambiguous resolution for *Gyrodactylus* taxonomy; however, the delimitation of closely related species in the genus is frequently controversial.

## 1.2 GYRODACTYLIDS ON EUROPEAN SALMONIDS

### 1.2.1 *GYRODACTYLUS SALARIS* AND *G. THYMALLI*

*Gyrodactylus salaris* Malmberg, 1957, is a notifiable pathogen of wild and cultured salmonids. Originally described from Atlantic salmon (*Salmo salar* L., 1758) the species potentially infects a wider range of salmonids hosts (Harris et al. 2004), including Arctic charr (*Salvelinus alpinus* L., 1758; e.g. Robertsen et al. (2007) and Harris et al. (2011)) and rainbow trout (*Oncorhynchus mykiss* Walbaum, 1792; e.g. Paladini et al. (2009)). *G. salaris* became known as the "salmon killer" after causing severe salmon epidemics in Norwegian watercourses in the 1970s (Johnsen 1978) and since then the parasite has caused severe salmon mortality in many Norwegian watercourses and a Russian river (Keret) draining into the White Sea. A number of Norwegian rivers have been treated with the indiscriminate poison rotenone in order to eradicate the parasite by eliminating the host (and causing

collateral damage of unpredictable dimension to the ecosystem). However, in the long term even this aggressive eradication policy seems futile as *G. salaris* re-occurred in several of the treated rivers. An excellent review on the history of the epidemics is provided by Bakke and co-workers (Bakke et al. 2007). Notably, the parasite is considered relatively harmless for salmon stocks in rivers draining into the Baltic sea (Bakke et al. 2002). The generally accepted conclusion is that the parasite was introduced from the Baltic region into East Atlantic salmon stocks, which lacked endogenous resistance to the parasite (Bakke et al. 2007). At least three independent introductions into Norway have been inferred based on molecular COI evidence (Hansen et al. 2003).

*Gyrodactylus thymalli* Zitnan, 1960 is an apparently benign parasite of European grayling (*Thymallus thymallus* L., 1758). Potential conspecificity of *G. salaris* and *G. thymalli* has been suggested initially based on morphological similarity (Malmberg and Malmberg 1993). In an early attempt to discriminate *Gyrodactylus* species based on molecular genetic approaches, Cunningham and co-workers failed to distinguish the two species genetically (Cunningham et al. 1995b). Later the ribosomal ITS regions were also shown to be invariable between *G. salaris* and *G. thymalli* across a wide geographical range (Cunningham 1997; Matejusova et al. 2001; Zietara and Lumme 2002) and potential conspecificity was again emphasized (Zietara and Lumme 2002). The ribosomal intergenic spacer region (IGS) was initially reported to differentiate between *G. salaris* and *G. thymalli*, with respect to the number, sequence and order of repeat motifs (Sterud et al. 2002), However, a re-evaluation following extended geographical sampling rejected the initially observed differences as non-diagnostic for species discrimination (Hansen et al. 2006). Not unexpectedly, the more rapidly evolving mitochondrial COI gene allowed for higher resolution and haplotypes of *G. salaris* and *G. thymalli* clustered into several distinct clades (haplogroups) mapping to host species (Hansen et al. 2003; Meinila et al. 2004). However, the clades were arranged in a star-like topology with weak statistical support at the basal nodes. Three taxonomic hypotheses have been put forward to explain the observed relationships as follows: *G. salaris* and *G. thymalli* may represent (i) two polytypic species, (ii) one polytypic species, or (iii) a complex of several sibling species (Hansen et al. 2003). Extended geographical sampling revealed further distinct lineages for *G. thymalli*, corresponding to major catchment areas (Hansen et al. 2007), but failed to improve the resolution of the basal relationships. Monophyly has thus never been unequivocally supported for either species. *G. salaris* is thought to originate from a host switch from European grayling onto Atlantic salmon (Meinila et al. 2004) and it was later

speculated that this host shift was preceded by a single hybridization event between two distinct lineages of *G. thymalli*, which was dated to the Eemian interglacial approximately 130,000 years ago (Kuusela et al. 2007). European grayling was thus assumed to be the plesiomorphic (ancestral) host for the complex. However, grayling exhibits a complex European phylogeographic history, probably predating the Pleistocene glacial cycles (Koskinen et al. 2000; Weiss et al. 2002; Gum et al. 2005). To formulate a conclusive phylogeographic hypothesis for *G. thymalli* geographic sampling must be extended across the entire natural distribution of *T. thymallus*. Currently, no data are available for a number of major European river systems, including the Elbe, Rhine, Danube, Loire and the Rhone, all of which support rich inter- and intrabasin genetic diversity of European grayling (Weiss et al. 2002; Gum et al. 2005).

One haplotype of *G. salaris* (haplotype F sensu Hansen et al. (2003), haplotype II sensu Meinila et al. (2004), later referred to as the rainbow trout clade by Kuusela et al. (2007)) shows an unexpectedly large geographic distribution. It has been reported from salmon in Norwegian and Russian rivers (Hansen et al. 2003; Meinila et al. 2004), as well as from Arctic charr in Norway (Olstad et al. 2007a; Robertsen et al. 2007), and on rainbow trout from a wide range of European fish farms (Meinila et al. 2004; Rokicka et al. 2007; Paladini et al. 2009). It is particularly interesting that the parasites isolated from Arctic charr in Lake Pålsbufjorden appeared to be non-pathogenic to Atlantic salmon (Olstad et al. 2007a), despite carrying the exact same mitochondrial haplotype as discussed above (Robertsen et al. 2007). This result was thoroughly re-evaluated and confirmed in a recent study (Ramirez et al. 2014). The ITS sequence identified for the Lake Pålsbufjorden population of *G. salaris* differed at one nucleotide position from the ITS sequence previously reported from *G. salaris* and *G. thymalli* populations across its entire range (Robertsen et al. 2007).

Despite the morphological similarity and the lack of clear genetic differentiation, experimental studies have demonstrated biological differences between *G. salaris* and *G. thymalli*. European grayling was interpreted as unable to sustain *G. salaris* in nature by Soleng and Bakke (2001). The tested parasite populations survived on grayling only ~35 days. An independent study concluded that Atlantic salmon is even less suitable as a host for *G. thymalli* than grayling is for *G. salaris* (Bakke et al. 2002). These results have been interpreted as evidence against conspecificity of *G. salaris* and *G. thymalli*. Kuusela and co-workers (2009) extensively sampled *G. salaris*/*thymalli* in the river Tornio. Despite a large sample size (n>300) and the sympatric occurrence of salmon and grayling in all sampling

sites, the study did not find a single incident of parasites with *G. salaris* mt haplotypes to occur on grayling or *G. thymalli* mt haplotypes to occur on salmon. This confirms the experimental evidence for the pronounced host specificity in a natural setting, which may be interpreted as mechanism for reproductive isolation between *G. salaris* and *G. thymalli* and thus as further evidence against conspecificity.

To date, *G. salaris* and *G. thymalli* are formally still considered distinct species by most authors. If a formal synonymization were undertaken, the valid name for the taxon would become *G. salaris* (*G. thymalli* would be the junior synonym) and a revision of the currently applied legislative frameworks, with respect to the prevention of the spread of *G. salaris*, would be required. An exhaustive review of the issue has been compiled by Bakke and co-workers (Bakke et al. 2007). The species are treated as *de facto* synonymized by a few authors (see e.g. Meinila et al. (2004) and Kuusela et al. (2009)) and a formal synonymization is currently being reviewed (Fromm et al. under review). Whatever the taxonomic status, the *G. salaris/thymalli* system offers unique opportunities for studying evolution in real time. Host switching is believed to represent the major driving force of speciation in gyrodactylids (see section 1.3 below). Second and third generation sequencing technologies now enable comprehensive genome-wide screening for "speciation genes", i.e. genes that retain reproductive isolation (Wu and Ting 2004) after potentially rapid evolution following the initial utilization of a new host species. Genome-wide data are currently available only for parasitic flatworms with immediate relevance for human health, such as schistosomes and tapeworms (Berriman et al. 2009; Consortium 2009; Tsai et al. 2013; Zheng et al. 2013). A first reference genome for *Gyrodactylus* will provide a paramount foundation for investigations of the genomic basis of host specificity and pathogenicity, topics relevant not only for gyrodactylid monogeneans.

### 1.2.2 *GYRODACTYLUS TEUCHIS* AND *G. TRUTTAE*

*G. teuchis* Lautraite, Blanc, Thiery, Daniel & Vigneulle, 1999 was originally described from rainbow trout from France (Lautraite et al. 1999), but the species seems to be rather generalistic and has since then been reported from a wider host range, including wild Atlantic salmon (Lautraite et al. 1999), brown trout (*Salmo trutta* L., 1758) and brook trout (*Salvelinus fontinalis* Mitchill, 1814) (Matejusova et al. 2001). *G. teuchis* is considered a cryptic species,

with strong morphological resemblance to *G. salaris* (Lautraite et al. 1999; Cunningham et al. 2001) (see Figure 2). The additional overlap in host range complicates discrimination of the two species and reliable identification is largely dependent on the use of molecular diagnostics. However, ITS sequences reveal *G. teuchis* to be distinct from *G. salaris* with a K2P distance of ~9% between the species. *G. teuchis* was reported from wild salmon and trout populations along the European Atlantic coast (Lautraite et al. 1999; Cunningham et al. 2001; von Gersdorff Jorgensen et al. 2008). Based on this distribution it was speculated that the species could originate either from an Iberian refugium or from the Channel River system (Bakke et al. 2007), However, the species wide host range and in particular its affinity to rainbow trout has likely enhanced its European spread through aquaculture and it was reported also from fish farms in Czech republic (Matejusova et al. 2001), Poland (Rokicka et al. 2007), Germany (Dzika et al. 2009) and Italy (Paladini et al. 2009). To date, *G. teuchis* has not been reported from wild salmonids in the UK or Fennoscandia (i.e. Norway, Finland and Sweden), and in particular its apparent absence from the UK appears to contradict the hypothesis of its Channel river origin.

*Gyrodactylus truttae* Gläser, 1974 has been described from the type host *Salmo trutta* and occurs south of the Baltic (Poland, Denmark, Germany, Czech Republic, Slovakia; Bakke et al. (2007)) and in the UK (Scotland and Ireland; Cunningham et al. (1995a)), but has not yet been recorded from Fennoscandia. It is considered a morphologically distinct non-pathogenic parasite of salmonids. *Salmo trutta*, *S. salar* and *Oncorhynchus mykiss* have been listed as potential hosts (Harris et al. 2004). As for *G. teuchis* anthropogenic spread of the parasite via aquaculture is likely. As of today, no hypotheses about the glacial history of *G. truttae* in Europe have been put forward.

Research on gyrodactylids infecting European salmonids is clearly biased towards the economically important *G. salaris*, yet, *G. teuchis* and *G. truttae* represent rich resources for the formulation of interesting evolutionary questions in their own rights. The ecological overlap with *G. salaris* in terms of host range also makes them highly relevant alternative models for addressing some of the unresolved questions surrounding the latter species. However, a thorough initial characterization of the morphological, ecological and genetic intraspecific diversity of *G. teuchis* and *G. truttae*, respectively, is a necessary prerequisite.

## 1.3 HOST SPECIFICITY AND SPECIATION IN *GYRODACTYLUS*

Gyrodactylids exhibit the widest host range of any monogenean flatworms. According to Bakke and co-workers the currently described ~ 400 *Gyrodactylus* species infect 20 of the 45 orders of fish (Bakke et al. 2002). On the species level, gyrodacylids were traditionally considered highly host specific, but according to a careful re-evaluation only about 30% of *Gyrodactylus* species occurred on just one single host and the previously presumed generally narrow specificity in the group was dismissed as an artefact based on the numerous descriptions of species collected only once (Bakke et al. 1992). Gyrodactylid host specificity is dictated by a number of mechanisms, which were previously classified into (i) ecological and behavioral mechanisms (host localization, recognition and attachment) and (ii) physiological mechanisms (establishment, growth and reproduction; see Bakke et al. (2002) for details). Host specificity in gyrodacylids is usually quantified based on factors such as reproductive rate and population growth of a given parasite species on a range of potential hosts. However, thorough experimental assessment is large limited to *G. salaris* (see Bakke et al. (2002) for a review).

Two basic mechanisms of speciation can be recognized for parasites, i.e. co-speciation and host switching. Viviparous gyrodactylids engage in a direct life cycle and lack specialized transmission strategies, they are, in other words, constantly facing opportunities to move between hosts (Kearn 1994). Bakke et al. (2007) made the case that most apparently similar species infecting the same or closely related hosts are frequently quite distantly related, while the most closely related species pairs tend to infect rather distantly related hosts (Bakke et al. 2007). Host switching rather than co-speciation is thus commonly considered the dominant mechanism of speciation in *Gyrodactylus*. However, co-speciation of gyrodactylids has also been demonstrated occasionally. Huyse and Volckaert (2005) investigated the evolutionary associations between European gobies (Gobiidae) and their *Gyrodactylus* parasites. They concluded that the radiation of highly host specific gill parasites was largely driven by host switching from the three-spined stickleback (*Gasterosteus arcuatus*) onto two gobiid genera *Pomatoschistus and Gobius* , followed by phylogenetically conserved host switching among various goby hosts (Huyse and Volckaert 2005). They suggested co-speciation for the less host specific fin-parasites resulting in several host-associated species complexes (Huyse and Volckaert 2005). Another case of possible host-parasite co-speciation has been reported for the Central American guppies *Poecilia reticulate* and *P. picta* and their respective

gyrodactylid parasites *G. turnbulli* and *G. pictae* (Cable et al. 2005). The mechanism of speciation is thought to have fundamental consequences for host specificity. Parasites which co-evolve, and potentially co-speciate, with their respective hosts are likely to evolve high host specificity, while species emerging from host switching events are likely to be less narrowly host specific (Bakke et al. 2002). However, host specificity is not necessarily a prerequisite for co-speciation (Huyse and Volckaert 2005). As species evolve over time, generalism, like host specificity, is sometimes a pure transitional state rather than a fixed trait (Nosil 2002).

## 1.4 INTERRELATIONSHIPS OF THE NEODERMATA

The obligate parasitic flatworms currently comprise three major groups (reviewed in Olson and Tkach (2005)): the Monogenea (Monopisthocotylea and Polyopisthocotylea), the Cestoda (tapeworms; Eucestoda and Cestodaria), and the Trematoda (flukes; Aspidogastrea and Digenea). The three groups are united as the Neodermata by the name giving Neodermis, i.e. a larval secondary syncytial tegument, which is believed to represent the key innovation of the group allowing for their immense radiation (Littlewood 2006). The Neodermata constitute one of the three largest groups of metazoan parasites of vertebrates (the other two being nematodes and arthropods), and include many species of medical and veterinary significance. The most prominent examples include *Schistosoma* spp., responsible for about 300,000 human deaths annually (van der Werf et al. 2003). The monophyly of the Neodermata is largely accepted, but the interrelationships within the group are still under discussion. The principal competing hypotheses are: (i) a sister group relationship between the Monogenea and Cestoda, based on putative morphological similarities, formulated in the "Cercomer-theory" (Janicki 1920), and supported by early molecular studies using 18S ribosomal DNA markers (Littlewood et al. 1999), and (ii) a sister group relationship between the Cestoda and Trematoda, more recently supported by nucleotide sequences of the combined 18S and 28S ribosomal genes (Lockyer et al. 2003), mitochondrial DNA (Park et al. 2007; Perkins et al. 2010), and microRNA loci (Fromm et al. 2013). Resolving this controversy is a necessary prerequisite for understanding the evolution of parasitism in the group. The Monogenea, including *Gyrodactylus*, are characterized by a direct ectoparasitic life style, while the Cestoda and Trematoda engage in endoparasitic life cycles varying in complexity, but usually involving at least one vertebrate and one invertebrate host. Fundamental questions include the plesiomorphic (ancestral) mode of parasitism, i.e. ecto- vs. endoparasitism first, and the plesiomorphic host, i.e. vertebrate vs. invertebrate hosts first, of the group. Genome wide data are currently only available for some important human parasites, such as *Schistosoma mansoni*

(Berriman et al. 2009) and *S. japonicum* (Consortium 2009), and the tapeworms *Echinococcus multilocularis, E. granulosus* (Zheng et al. 2013), *Taenia solium* and *Hymenolepis microstoma* (Tsai et al. 2013). In a recent, highly significant study Tsai and co-workers presented several genomic traits interpreted as adaptations to the parasitic lifestyle in flatworms, based on previously and newly sequenced genomes (Tsai et al. 2013). However, no monogenean is currently available for comparative genomics and phylogenomics approaches, representing a major limitation for a comprehensive evaluation of the issue.

## 1.5 OBJECTIVES OF THE PRESENT THESIS

The PhD fellowship summarized in the current thesis has been based around the biology and evolution of ecotparasitic gyrodactylid flatworms. Some parts of the presented work are largely descriptive, providing the foundation for future studies, while other parts address fundamental questions. A brief outline of the objectives of the current project and the goals of the individual papers are given below.

*G. teuchis* is widely spread across Europe and, as outlined in section 1.2.2 above, bears a striking morphological similarity to *G. salaris.* Furthermore, it exhibits a marked overlap in host range with the latter species, which led to the conclusion of *G. teuchis* being a cryptic species. The first description of *G. teuchis* utilized samples from two geographically well separated French populations of brown trout (Brittany and Western Pyrenees), providing valuable insight into the overall range of morphological variation (Lautraite et al. 1999). However the study did not address morphological variation within the populations. Subsequently, a re-description was published providing morphological data from a *G. teuchis* population on farmed rainbow trout (Cunningham et al. 2001). In recent years morphological species identification in gyrodactylids has made increasing use of geometric morphometrics, and multivariate approaches, such as Principal Component Analysis (PCA), which has been applied to distinguished between morphologically very similar species, in particular *G. salaris* and *G. thymalli* (Shinn et al. 2004; Olstad et al. 2007b). The morphological characterization of *G. teuchis* provided in the initial descriptions were far from comprehensive (Lautraite et al. 1999; Cunningham et al. 2001) and as a prerequisite for comparative multivariate analyses with its morphologically most similar congeners it was necessary to establish the relevant set of morphometric measurements also for *G. teuchis*. Suitable parasite material from a natural

Austrian brown trout population was at hand to achieve this goal. At the same time the material would enable a first assessment of morphological and genetic variation in a natural population of a Danube tributary and allow for comparisons with the previous studies (albeit limited to the morphological characters presented therein).

In recent years it has become feasible to utilize entire mitochondrial (mt) genomes for addressing questions in evolutionary biology, population genetics, as well as in phylogeographic and phylogenetic studies in animals. The mt genomes of *G. salaris* (Huyse et al. 2007), *G. thymalli* (Plaisance et al. 2007) and *G. derjavinoides* (Huyse et al. 2008) have been published previously. Another goal to be achieved as part of the current PhD fellowship was to complement the available mt genomes with two additional species, i.e. *G. teuchis* and *G. truttae*, and to investigate the mode of evolution of mitochondrial genes/genomes in gyrodactylids. To allow for inter- and intraspecific comparisons individuals from several populations for each species were to be included. Until recently, the most commonly used method to characterize the nucleotide sequence of entire mt genomes involved the initial amplification of the mt genome, represented by only a small number of long overlapping DNA fragments, by long range PCR. These were subsequently sequenced via primer walking. However, the design of suitable long range PCR primers relies on mitochondrial reference genomes of relatively closely related organisms, which might not be available for non-model organisms. Furthermore, the traditional long range PCR/primer walking approach may require large amounts of man-hours (C. Hahn, personal observation). The initial goal was thus to design an optimized bioinformatics pipeline for the assembly of mt genomes directly from genomic DNA derived short NGS reads.

The throughput of NGS platforms has increased dramatically during the past years, along with a marked reduction of per base cost. Genome sequencing projects, especially for organisms with small to medium range genome size, have become increasingly feasible for individual labs without the need for large cooperative consortia. A first reference genome for a monogenean flatworm would at last allow the evaluation of the disputed relationships within the obligate parasitic Neodermata (see section 1.4 above) using a phylogenomics approach. Furthermore, a reference genome obtained from a pathogenic strain of *Gyrodactylus salaris* would provide an immensely valuable milestone towards understanding the genomic bases of pathogenicity and host specificity (see section 1.2.1 above) in gyrodactylids and flatworms in general. Last but not least, characterizing the gene complement of *G. salaris* will be of major significance for the development of drug based control strategies, specifically targeting

essential biochemical pathways of the parasite instead of the ecologically unsustainable eradication policy currently implemented by the Norwegian government. The *G. salaris* genome project was thus launched by Lutz Bachmann and as part of the current PhD project a high quality draft genome was to be constructed.

Studies targeting the phylogeographic history of gyrodactylid flatworms are scarce. A few notable exceptions exist for *G. salaris* and *G. thymalli* (Hansen et al. 2003; Meinila et al. 2004; Hansen et al. 2007; Kuusela et al. 2007), but, as pointed out in section 1.2.1 above, due to the lack of resolution and/or the limited geographical sampling the presented results are far from conclusive. Another objective of the current PhD was thus to improve the existing phylogeographic hypotheses for *G. thymalli* and add novel information for other selected salmonid infecting gyrodactylids, i.e. *G. teuchis* and *G. truttae* (see section 1.2.2 above). The first prerequisite was to extend the existing geographic sampling to previously neglected river basins, e.g. the Loire, Rhone, Danube, Elbe and Rhine systems, which all harbour native salmonids. Samples were then to be analysed using standard molecular markers, such as COI, to build on existing datasets. Results were to be integrated with existing phylogeographic hypotheses for the hosts, i.e. *T. thymallus* (Koskinen et al. 2000; Weiss et al. 2002; Gum et al. 2005) and *S. trutta* (Bernatchez 2001; Weiss et al. 2001; Cortey and Garcia-Marin 2002; Cortey et al. 2004; Cortey et al. 2009; Lerceteau-Kohler et al. 2013).

# 2. METHODOLOGICAL CHALLENGES

## 2.1 *DE NOVO* GENOME ASSEMBLY OF NON-MODEL ORGANISMS

In recent years advances in sequencing technology have made the characterization of entire genomes feasible. In particular the emergence of so-called Next Generation Sequencing (NGS) platforms has enabled the generation of unprecedented amounts of molecular sequence data (Metzker 2010). The reconstruction of genomes from shorter fragments (often termed reads) into long contiguous fragments (contigs) without any reference information, commonly termed *de novo* assembly, has become an increasingly dynamic field of research, and new assembly algorithms are continuously being developed, assessed and improved (Miller et al. 2010; Earl et al. 2011; Nagarajan and Pop 2013). However, the only valid conclusion from the discussion is that there is currently no consensus as to which is the "best" assembly algorithm is, but rather that the choice is heavily dependent on the peculiarities of the dataset/genome in question (Earl et al. 2011; Salzberg et al. 2012; Bradnam et al. 2013). Nevertheless, genome assembly for eukaryotes with small- to medium-size genomes has become a democratized bottom up enterprise (Kumar et al. 2012). The initial major challenge for small unculturable invertebrates remains the relatively high amount of genomic DNA required for successful NGS library preparation, which often exceeds DNA quantities that can be extracted from single individuals. Once sufficient data are generated the actual assembly process can be computationally quite expensive, i.e. might require hundreds of thousands of CPU hours and considerable amounts of RAM. The latest assembly for *Drosophila melanogaster* using the Celera Assembler (CA) PBcR pipeline (Koren et al. 2012) required >600,000 CPU hours (step by step guide through the assembly including relevant stats can be found at http://cbcb.umd.edu/software/pbcr/dmel.html, accessed 04/2014). However, at the same time as High Performance Computing Clusters (HPCCs) are getting more and more powerful (and increasingly accessible to researchers) new algorithms are being explored to reduce the computational footprint, while keeping the quality impairment of the assembly to an acceptable minimum. One of the most notable recent innovations was digital-normalization (diginorm) of NGS data prior to genome assembly (Brown et al. 2012; Pell et al. 2012), which can be used to remove data redundancy theoretically without loss of information. Currently a new class of memory-efficient genome assemblers is emerging, which are capable of generating draft assemblies of reasonable quality on conventional multi-purpose computers

(Kleftogiannis et al. 2013). The Minia assembler, for example, performed a complete *de novo* assembly of human genome short reads using < 6 GB of memory in < 1 day (Chikhi and Rizk 2012). Clearly, with sequencing costs steadily decreasing, there currently exists a tradeoff between assembly quality and computational footprint. In my opinion, the resources and effort devoted to a particular genome assembly should be tailored to the biological question which is being tackled. The highly memory-efficient assemblers are unlikely to yield a chromosome level genome assembly, but the resulting representation might well be more than sufficient for e.g. the prediction of initial gene models. A commonly used metric to assess genome assemblies is the so-called N50. The Assemblathon 1 paper (Earl et al. 2011) defines N50 as ".. a weighted median of the lengths of the sequences it contains, equal to the length of the longest sequence s, such that the sum of the lengths of sequences greater than or equal in length to s is greater than or equal to half the length of the genome being assembled." An assembly N50 of around the median gene length of an organism is sufficient to expect about 50 % of all genes to be uninterruptedly present in a draft (Yandell and Ence 2012). According to Flybase (http://flybase.org/) the median gene size of *Drosophila* is 1,560 bp (Daines et al. 2011). The N50 of the latest assembly for *D. melanogaster* is > 15,000,000 bp (http://cbcb.umd.edu/software/pbcr/dmel.html, accessed 04/2014).

An excellent review on *de novo* genome assembly has been recently compiled by Nagarajan and Pop (2013). *De novo* genome assembly relies on the assumption that highly similar DNA fragments originate from the same position within the genome (Nagarajan and Pop 2013). Genomic repeats violate this assumption, as they may yield highly similar sequences that originate from different places in the genome (Nagarajan and Pop 2013). The complexity of a genome assembly and the chances of resolving it are dictated by the relationships between read length provided by the NGS sequencing platform of choice and the length of repeats in the DNA being assembled. Consequently *de novo* assemblies might be classified along a continuum from trivial (if all repeats are shorter than the read length) to impossible (if the information contained within the reads is insufficient to identify the correct sequence reconstruction from an exponential number of equally good alternatives) (Nagarajan and Pop 2013). A major technological advance in the field was the recent release of the PacBio RS sequencing instrument, the only long-read, single-molecule sequencer currently available. PacBio utilizes a novel approach to observe DNA polymerization in real time (Eid et al. 2009) and the method has been successfully used to sequence recent outbreak genomes (Chin et al. 2011; Rasko et al. 2011). However, the inherently high error rate of ~ 18 % (Nagarajan and

Pop 2013) required the development of efficient error correction algorithms. Initially the use of highly accurate short read data produced by the Illumina platform was proposed as a promising strategy (Koren et al. 2012) but most recently the hierarchical genome-assembly process (HGAP), relying exclusively on data produced by the PacBio instrument, has been introduced (Chin et al. 2013). The PacBio RS system produces reads of variable length and the typical read length frequency distributions peak at relatively short lengths of a few kb (note that the term *short* in respect to reads is usually associated with Illumina style read lengths of a few hundred bp), but also contains substantially longer reads of currently up to ~ 30 kb in low frequency. Short reads represent consensus reads produced from reading the exact same DNA molecule several times and are thus of high accuracy (Steve Picton, Pacific Bioscience Europe, personal communication). The HGAP method makes efficient use of the shorter reads to correct the longest reads. Only these long corrected reads are subsequently used in the assembly process, substantially reducing its complexity (Chin et al. 2013). The HGAP approach has been implemented in the latest version of the Celera assembler (v.8.1) PBcR pipeline (Koren et al. 2012) and recent studies suggests that the majority of known bacterial and archaeal genomes can be assembled without gaps, at finished grade quality, using a single PacBio RS sequencing library and the hierarchical assembly approach (Koren et al. 2013).

Another challenge in genome assembly projects is the potential contamination of the genomic DNA extracts with non-target DNA. In particular for projects focusing on small invertebrates, contamination may be unavoidable due to the presence of e.g. (i) symbiotic organisms attached to or within the sampled specimens or (ii) ingested food. The latter type of contamination is probably of reduced significance if the genomic library is sequenced on a long-read instrument (see above), but short-read technologies such as Illumina might readily pick up degraded DNA fragments. Recently, an *in silico* approach to aid partitioning of draft assembly contigs, and the reads that contribute to these contigs, has been presented (Kumar et al. 2013). The presence of genomic DNA of a multitude of organisms in a sample is a particular challenge in shotgun metagenome projects, which specifically aim at capturing potentially all genomes present in an environment. In addition to the exponentially increasing complexity of such datasets, an inherent difficulty for shotgun metagenome projects is that some organisms may simply be less abundant in a given environmental sample than others. In order to get a glimpse on low abundance organisms it may thus be necessary to produce huge sequencing depths, which in turn will increase the required computational resources. The

removal of data redundancy as implemented in the diginorm approach (Pell et al. 2012) (see also above) is thus essential for such projects.

## 2.2 EXTENDING GEOGRAPHIC SAMPLING FOR PHYLOGEOGRAPHIC STUDIES

Extending geographic sampling by the scale aimed at in the current project requires a large collaborative effort. Non-pathogenic gyrodactylids on salmonid fish are not necessarily highly abundant (C. Hahn, personal observation) and the most effective method for obtaining the potentially large numbers of host fish necessary is electro-fishing. Hosts can subsequently be screened for parasites alive. The legislative framework for applying the electro-fishing methodology for biological sampling differs between European countries. Luckily, ongoing European wide assessments of the ecological status of freshwater systems in the course of the EU Water Framework Directive (http://ec.europa.eu/environment/water/water-framework/index_en.html) provided a unique opportunity for obtaining gyrodactylid samples. The support of local research teams and authorities was paramount for the current project.

# 3. RESULTS AND BRIEF DISCUSSION

## PAPER I

The first paper of the current thesis focusses on *G. teuchis*, a common non-pathogenic parasite of European salmonids. The species bears striking morphological and ecological (in terms of host range) similarity to *G. salaris* and is widely accepted as principal example of a cryptic *Gyrodactylus* species. *G. teuchis* has been described from the European Atlantic basin. We reported *G. teuchis* for the first time from the Danube system (Black sea basin). The local host population (brown trout) is presumably unaffected by stocking and has never been in contact with rainbow trout. The described population of *G. teuchis* is thus likely to represent a natural occurrence in the Danbue system, instead of having been locally introduced via the vector rainbow trout. The study demonstrated that the so-called cryptic species *G. teuchis* can indeed be discriminated morphologically from *G. salaris* using multivariate Principal Component Analysis (PCA) based on a set of morphometric measurements. Interestingly, the morphological variation identified within this one population was comparable to the overall morphological variation thus far reported for European *G. teuchis*. Ribosomal ITS sequences obtained from the population were largely identical to the previously reported sequence. However, we also identified an intra-individual heterogeneity of ITS 1 which occurred in > 60 % of the analysed individuals (n=31). The observation was confirmed by using the restriction enzyme BclI to selectively digest one of the variants and interpreted as evidence for a non-fully homogenized rDNA locus. The paper reviews previously published information on the species and summarizes it in a concise, formal re-description as a foundation for future studies. Furthermore, the paper provides the first high resolution images of the opisthaptoral hard parts for the species, produced using a new modified preparation protocol relying on polylysine coated slides for improved hook adherence. The protocol has been adopted successfully for *Diplozoon* spp. (Monogenea; Q. Dos Santos, University of Johannesburg, unpublished).

## PAPER II

Paper II presents a novel *in silico* approach for the reconstruction of mitochondrial (mt) genomes on non-model organisms directly from total genomic DNA derived NGS data. No prior enrichment or amplification of mt DNA is needed. The new approach requires no reference genome from the species in question and is thus highly applicable to non-model organisms. The approach is straightforward even if only (i) distantly related mitochondrial genomes or (ii) mitochondrial barcode sequences are available as starting-reference sequences or seeds, respectively. The principal functionality of the new method, dubbed MITObim (mitochondrial baiting and iterative mapping), was demonstrated on real NGS data generated from genomic DNA extracts of the monogenean ectoparasites *G. thymalli* and *G. derjavinoides*. MITObim assembled highly accurate representations (>99.5 % base accuracy) of the mt genomes of both species in less than 24 hours on a standard desktop computer, demonstrating its supremacy over existing tools in terms of accuracy, runtime and memory footprint. Parasitic DNA extracts are inevitably contaminated with host DNA. In an additional case study we thus used MITObim to reconstruct the mt genomes of the respective salmonid hosts of the two parasite species from the genomic DNA extracts. Another case study based on simulated data demonstrates the high specificity of the approach and its potential applicability to metagenomic or pooled sequence data. The pipeline is written in Perl and facilitates modules of the MIRA assembly suite (http://sourceforge.net/projects/mira-assembler/). The pipeline is tailored for the use by biologists with modest bioinformatics skills. The scripts, including a step by step tutorial, were made freely available (https://github.com/chrishah/MITObim), and the pipeline is constantly being developed and improved. The method has received 17 citations in peer-reviewed journals in its first year after publication in *NAR* (based on PubMed, 05/2014). The approach has been adopted in a recently published ultra-fast mitogenome recovery pipeline (Gan et al. 2014) and one publication has even linked our method to “a revolution for the sequencing of organellar genomes“ (Besnard et al. 2014). Although originally designed for the use with animal mitochondria, another recent study has demonstrated the methods applicability for the assembly of substantially larger chloroplast genomes (Mariac et al. 2014). In house we have used the pipeline to assemble novel mitochondrial genomes for *G. teuchis* and *G. truttae* and the corresponding manuscript, investigating intra- and interspecific modes of evolution in gyordactylid mitochondrial genomes, is anticipated for submission later in 2014. MITObim

has furthermore been used to reconstruct complete mt genomes from DNA extracts obtained from ancient Bowhead whale remains in a project focussing on the Paleo-population genomics of the Spitzbergen stock of Bowhead whales (Bachmann, Hahn et al. manuscript in preparation).

## PAPER III

Paper III represents a major milestone for understanding the evolution of parasitism in flatworms. It presents the first draft genome of a monogenean flatworm, the economically important ectoparasite *G. salaris*. The ectoparasitic Monogenea comprise a major part of the obligate parasitic flatworm diversity. We resolve the disputed phylogenetic relationships of the three major parasitic groups in the obligate parasitic Neodermata, i.e. Monogenea, flukes (Trematoda) and tapeworms (Cestoda), using a large scale phylogenomic approach utilizing a total of over 1,700 genes (> 500,000 amino acid positions). The Monogenea were found basal to the endoparasitic tapeworms and flukes, which implies ectoparasitism being the plesiomorphic (ancestral) mode of parasitism in the group. Putative genomic adaptations to parasitism have been investigated before in the endoparasitic groups for species of major human health concern, such as the blood fluke *Schistosoma mansoni* and the dog tapeworm *Echinococcus granulosus* (Tsai et al. 2013). With the first monogenean genome at hand previous hypotheses were re-evaluated in a comparative genomic approach using the genomes of seven parasitic flatworm species. We identified a number of shared genomic features for the ecto- and endoparasitic lineages, such as a substantial reduction of the core bilaterian gene complement, including homeodomain-containing genes, and a loss of the *piwi* and *vasa* genes, which are considered essential for animal development. Furthermore we inferred the shared loss of functional fatty acid biosynthesis pathways and the absence of peroxisomes, the latter organelles presumed ubiquitous in eukaryotes except for parasitic protozoans. The gene space in the draft genome was found to be > 95 % complete based on the presence of a set of conservative core genes. Overall, we identified a total of 15,488 gene models for *G. salaris* and were able to obtain putative functional annotations for 7,102 of these genes. The number of gene models found, as well as basic gene statistics (average gene length, average number of exons, etc) is largely concordant with findings for previously sequenced parasitic flatworms, which we interpret as an indication for the high quality of the presented genome and gene predictions.

## PAPER IV

The final study included in this thesis has focussed on the "neglected" salmonid-infecting species *G. teuchis* and *G. truttae*. The study is the first to investigate the distribution and phylogeographic history of these species on a broader European scale. Samples from most major European rivers systems (Danube, Elbe, Rhine, Loire, Rhone) were included in the study. *G. teuchis* appears to be the most abundant gyrodactylid on European salmonids and was detected in almost all sampled mainland European rivers from the Iberian Peninsula to the Balkans. *G. truttae* was generally less abundant and was not detected on the Iberian Peninsula and the Balkans. However, in contrast to *G. teuchis* it was detected in the UK. We observed fundamentally different intraspecific phylogenetic structure for the two species, based on the mitochondrial COI gene. *G. truttae* exhibits very low genetic diversity and no obvious phylogeographic structure across its European distribution. In contrast, *G. teuchis* is characterized by several deeply diverging lineages which largely correspond to the major European lineages of the hosts *Salmo* spp.. Significant global congruence between the host and parasite phylogenies was inferred using co-phylogenetic methods. Thus our results firmly support co-speciation for *G. teuchis* and its hosts. Co-speciation is a fundamental concept of evolutionary biology and intuitively appealing, yet in practice hard to prove as it is often blurred by other evolutionary processes. It was thus somewhat surprising to find a strong signal of co-speciation in a host-parasite system that is probably affected significantly by human mediated translocation of host stocks. The major split within *G. teuchis* appears to coincide with the initial divergence of brown trout and Atlantic salmon and was dated to ~ 1.5 My before present, using a Bayesian framework based on an indirect calibration point obtained from the host phylogeny. The presence of *G. teuchis* in Europe thus predates Pleistocene glaciations.

# 4. CONCLUSION AND FUTURE PERSPECTIVES

The current PhD thesis presents a number of significant contributions to the field of evolutionary biology in general and evolutionary parasitology in particular. It offers novel insights into mechanisms of evolution (**PAPER III**, **PAPER IV**), methodological advances (**PAPER II**) and overall important foundations for future research (**PAPER I**, **PAPER II**, **PAPER III**, **PAPER IV**).

A number of projects are currently ongoing that are based on data, results and/or methods presented as part of the current thesis. For example, for the purpose of studying the evolution of mitochondrial genomes in gyrodactylids, we generated low coverage Illumina data for a number of *Gyrodactylus* species. Subsets of the data have been used in different studies, such as **PAPER II** or as part of the Masters project "Phylogenetic footprint of microRNAs in *Gyrodactylus*" by S. Burow. Furthermore, the MITObim approach (**PAPER II**) has been used to reconstruct ~ 15 mitochondrial genomes from the NGS data obtained from a multiplexed Illumina run, including a number of novel genomes for *G. teuchis* and *G. truttae*. Combining our data with the published mitochondrial genomes of *G. derjavinoides* (Huyse et al. 2008), *G. salaris* (Huyse et al. 2007) and *G. thymalli* (Plaisance et al. 2007) and previously unpublished data from further strains of the latter two species (L. Bachmann et al., unpublished) results in an exciting dataset containing five species and a minimum of four mitogenomes per species from geographically isolated populations. The analyses targeting the mode of intra- and interspecific evolution of mt genomes in gyrodactylids are ongoing based on this dataset.

While the multiplexing approach used for generating the gyrodactylid NGS data resulted in an extensive representation of mitochondrial reads for the individual sample, the coverage of the nuclear genomes is substantially lower (< 10x). *De novo* sequencing projects usually aim at a coverage of at least 50x, but resequencing projects often target significantly lower coverage levels (Alex Buerkle and Gompert 2013). For a comparative approach across several species one might expect some allele dropout due to the low coverage and potential differences between species. However, mapping our low coverage data back to the reference genome of *G. salaris* (**PAPER III**) would be a straightforward approach and might well identify biologically relevant substitutions/indels differentiating the strains/species. With an appropriate sampling design such an approach could be used very efficiently to finally

identify the genomic bases of pathogenicity in *G. salaris* strains and shed light on the genomic mechanisms underlying host specificity.

Recently, first attempts towards population genetics using microsatellites have been made for gyrodactylids (Faria et al. 2011; Schelkle et al. 2012), but to date no population genomics methods have been applied. The small amounts of genomic DNA that can be extracted from single individuals are currently insufficient for most NGS library preparation protocols. However, a few kits exist which claim to require only minimal amounts of input DNA (<5ng). Initial tests should only require a relatively small investment and I hope to soon propose a small project to an appropriate funding body. Getting insights into the genomics of *Gyrodactylus* infections on a population level seems like an exciting prospect indeed. Combining this approach with carful experimental work holds immense potential to understanding the initial steps of speciation via host switching in gyrodactylids. The relatively generalistic species *G. teuchis* would be a highly relevant and ideal model system. It basically shares the host range of *G. salaris*, but is considered a non-pathogenic species with a wide natural European distribution (**PAPER IV**) and intraspecific (might be subject to discussion) mitochondrial diversity far exceeds the diversity reported from the *G. salaris*/*G. thymalli* complex. Hybridization experiments could be attempted along a gradient of phylogenetic relatedness to investigate potential post-zygotic isolation on a genome level. Experimental work with *G. salaris* is controlled by a strict legislative framework. As *G. teuchis* is a non-pathogenic and naturally widely distributed species on salmonids, and the regulations for the experimental setup might be expected relatively liberal in comparison to *G. salaris*, a notifiable pathogen often dubbed as "the salmon killer".

The extensive gyrodactylid sampling in European rivers systems conducted as part of this PhD fellowship has yielded a large collection of *Gyrodactylus* material. Samples of *G. teuchis* and *G .truttae* were incorporated in **PAPER IV**, and led to exciting insights into a rare case of co-speciation in gyrodactylids. However, we also substantially extended the geographic sampling for *G. thymalli* to previously neglected river systems, i.e. the Loire, Rhine, Rhone systems and the Adriatic basin, as well as a more in depth sampling of Austrian Danube system. Combining the new samples with previously published data (Hansen et al. 2003; Meinila et al. 2004; Hansen et al. 2007) reveals a number of new clades for the Adriatic, Danube, Loire, and Rhone catchments. However, the principal host *T. thymallus* exhibits a somewhat more complex history of colonization and admixture in European rivers than brown

trout (Koskinen et al. 2000; Weiss et al. 2002; Gum et al. 2005) and the analyses using the combined dataset are ongoing.

Another interesting approach to gyrodacylid phylogeography was explored in a further subproject of the current PhD project. Salmonids are particularly sensitive to human impacts, such as water pollution, overfishing and use of hydropower, which can significantly affect local fish populations and has frequently led to local extinction of formally abundant fish stocks. Historical changes in fish diversity are well documented in ichthyological collections, but so far no attempts have been made to explore the gyrodactylid ectoparasites unwittingly collected together with their hosts. We screened selected salmonids deposited in NHM Vienna, MNHN Paris and MfN Berlin (generously funded by the SYNTHESYS program: http://www.synthesys.info/) and were able to detect a number of *Gyrodactylus* individuals on the historic fish samples, most notably specimens from Atlantic salmon (collected 1876) and grayling (collected 1880, see Figure 3). The parasites were identified using hook morphology as *G. derjavinoides* and *G. thymalli*, respectively.

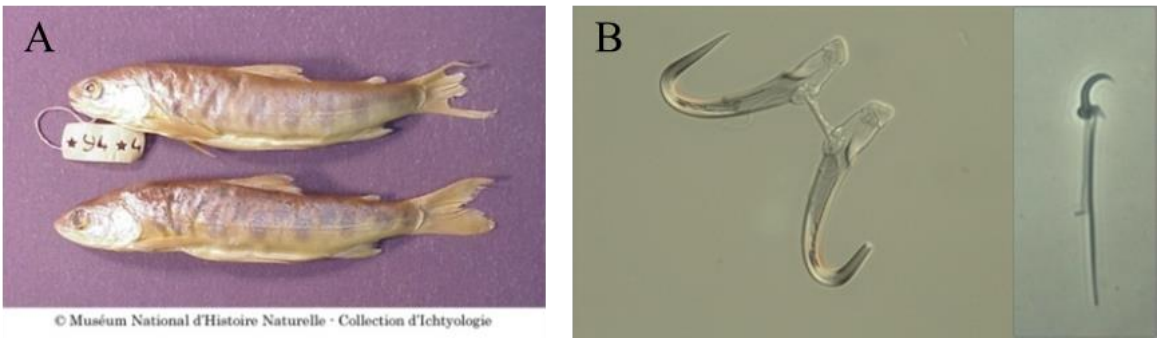



**Figure 3**: (A) Historical salmon samples from the ichthyological collection of the MNHN Paris and principal source for historical gyrodactylids. (B) Opisthaptoral hard parts of *G. thymalli* obtained from a historical sample of European grayling at the NHM Vienna (collected 1880, Danube tributary).

We have explored the suitability of this historical material for museomics approaches and managed to amplify parts of the mitochondrial COI gene for the latter specimen. More samples have subsequently been collected but the quality of the samples varies and the protocols for preparation and DNA extraction of this precious material require careful optimization, which was outside the scope of this thesis. However, our pilot study represents an important proof of concept and the collected material will be an exciting source of information for future projects.

One of the main findings of **PAPER III** has been a substantial loss of core bilaterian genes in the course of the evolution of parasitism in flatworms. For the paper we focused on a small number of genes/gene families previously emphasized in other studies (Tsai et al. 2013). The straightforward logical next step would be to identify overall functional patterns linked to the entirety of lost genes. Our results indicated that the currently predicted gene models are of high quality, but before attempting such analyses RNAseq data should be generated and used for in depth validation and potential improvement of the *G. salaris* gene models initially identified in **PAPER III**. Investigations into differential expression patterns between strains of *G. salaris* might also give highly relevant insights into gyrodactylid pathogenicity. Recently, the microRNA complement of *G. salaris* has been characterized. The study inferred a gradual loss of conserved microRNA families in the course of the evolution of Neoderamta (Fromm et al. 2013) which is reminiscent of the findings of general reduction in **PAPER III**. However, the authors also identified the emergence of novel microRNA families in the individual parasitic lineages (Fromm et al. 2013). Although the two studies were the result of a tightly linked collaboration, an integration of both datasets to identify the genomic targets regulated by the microRNA loci has been out of the scope of either project. A follow up study could easily explore this, but again RNAseq data should be obtained first, to verify the current *G. salaris* gene models.

The genome of *G. salaris*, the first monogenean to be sequenced, represents an important milestone towards understanding the evolution of parasitism in flatworms. **PAPER III** resolves the phylogenetic relationships of the major parasitic flatworm groups using a phylogenomics approach. However, a number of open questions remain that could not be tackled based on the existing dataset. With only one single representative of the Monogenea present in the analysis it was not possible to address the potential paraphyly of the group, with respect to the Monopisthocotylea and Polyopisthocotylea (Justine 1998; Perkins et al. 2010). Future studies should thus extend the taxon sampling within the Monogenea.

Based on our results we conclude that the current assembly for *G. salaris* represents a high quality draft and a key reference for numerous future studies. Additional sequencing data, such as mate pair libraries or long-read data generated by the PacBio RS instrument (see section 2.1 above), could certainly improve the quality and continuity of the current version. However, as stated above, we believe that the sequencing effort should be tailored to the biological questions that are being tackled. We are convinced that a number of questions (see

above) can readily be addressed based on the current version and an improvement of the draft genome should not have the highest priority at this stage.

## REFERENCES

Alex Buerkle C, Gompert Z 2013. Population genomics based on low coverage sequencing: how low should we go? Mol Ecol 22: 3028-3035. doi: 10.1111/mec.12105

Bakke TA, Cable J, Harris PD 2007. The biology of gyrodactylid monogeneans: the "Russian-doll killers". Adv Parasitol 64: 161-376. doi: 10.1016/S0065-308X(06)64003-7

Bakke TA, Harris PD, Cable J 2002. Host specificity dynamics: observations on gyrodactylid monogeneans. International Journal for Parasitology 32: 281-308.

Bakke TA, Harris PD, Jansen PA, Hansen LP 1992. Host specificity and dispersal strategy in gyrodactylid monogeneans, with particular reference to Gyrodactylus salaris Malmberg (Platyhelminthes, Monogenea). Diseases of Aquatic Organisms 13: 63-74.

Bernatchez L 2001. The evolutionary history of brown trout (Salmo trutta L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation. Evolution 55: 351-379.

Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama ST, Al-Lazikani B, Andrade LF, Ashton PD, et al. 2009. The genome of the blood fluke Schistosoma mansoni. Nature 460: 352-U365. doi: 10.1038/Nature08160

Besnard G, Jühling F, Chapios E, Zedane L, Lhuillier E, Mateille T, Bellafiore S 2014. Fast assembly of the mitochondrial genome of a plant parasitic nematode (Meloidogyne graminicola) using next generation sequencing. Comptes Rendus Biologies in press.

Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, et al. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. Gigascience 2: 10. doi: 10.1186/2047-217X-2-10

Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH 2012. A Reference-free algorithm for computational normalization of shotgun sequencing data. arXiv:1203.4802v2.

Cable J, Harris PD 2002. Gyrodactylid developmental biology: historical review, current status and future trends. International Journal for Parasitology 32: 255-280.

Cable J, van Oosterhout C, Barson N, Harris PD 2005. Gyrodactylus pictae n. sp. (Monogenea: Gyrodactylidae) from the Trinidadian swamp guppy Poecilia picta Regan, with a discussion on species of Gyrodactylus von Nordmann, 1832 and their poeciliid hosts. Systematic Parasitology 60: 159-164. doi: 10.1007/s11230-004-6348-4

Chikhi R, Rizk G 2012. Space-efficient and exact de Bruijn Graph representation based on a Bloom Filter. WABI.

Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10: 563-569. doi: 10.1038/nmeth.2474

Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, et al. 2011. The origin of the Haitian cholera outbreak strain. N Engl J Med 364: 33-42. doi: 10.1056/NEJMoa1012928

Cohen J. 1977. Reproduction. London, UK: Butterworth.

Consortium SjGSaFA 2009. The Schistosoma japonicum genome reveals features of host-parasite interplay. Nature 460: 345-351. doi: 10.1038/nature08140

Cortey M, Garcia-Marin JL 2002. Evidence for phylogeographically informative sequence variation in the mitochondrial control region of Atlantic brown trout. Journal of Fish Biology 60: 1058-1063. doi: 10.1006/jfbi.2002.1910

Cortey M, Pla C, Garcia-Marin JL 2004. Historical biogeography of Mediterranean trout. Molecular Phylogenetics and Evolution 33: 831-844. doi: 10.1016/j.ympev.2004.08.012

Cortey M, Vera M, Pla C, Garcia-Marin JL 2009. Northern and Southern expansions of Atlantic brown trout (Salmo trutta) populations during the Pleistocene. Biological Journal of the Linnean Society 97: 904-917.

Cunningham CO 1997. Species variation within the internal transcribed spacer (ITS) region of Gyrodactylus (Monogenea:Gyrodactylidae) ribosomal RNA genes. J Parasitol 83: 215-219.

Cunningham CO, McGillivray DM, MacKenzie K, Melvin WT 1995a. Discrimination between Gyrodactylus salaris, G. derjavini and G. truttae (Platyhelminthes: Monogenea) using restriction fragment length polymorphisms and an oligonucleotide probe within the small subunit ribosomal RNA gene. Parasitology 111 ( Pt 1): 87-94.

Cunningham CO, McGillivray DM, MacKenzie K, Melvin WT 1995b. Identification of Gyrodactylus (Monogenea) species parasitizing salmonid fish using DNA probes. Journal of Fish Biology 18: 539-544.

Cunningham CO, Mo TA, Collins CM, Buchmann K, Thiery R, Blanc G, Lautraite A 2001. Redescription of Gyrodactylus teuchis Lautraite, Blanc, Thiery, Daniel & Vigneulle, 1999 (Monogenea: Gyrodactylidae); a species identified by ribosomal RNA sequence. Systematic Parasitology 48: 141-150.

Daines B, Wang H, Wang L, Li Y, Han Y, Emmert D, Gelbart W, Wang X, Li W, Gibbs R, et al. 2011. The Drosophila melanogaster transcriptome by paired-end RNA sequencing. Genome Res 21: 315-324. doi: 10.1101/gr.107854.110

Dzika E, Maciejewska IW, Hoffmann RW, Oidtmann B 2009. The Gyrodactylidae fauna of rainbow trout Oncorhynchus mykiss Walbaum 1792 in the Rogg breeding pound in Bavaria, Germany. Parasitol Res 104: 671-676. doi: 10.1007/s00436-008-1244-5

Earl D, Bradnam K, St John J, Darling A, Lin D, Fass J, Yu HO, Buffalo V, Zerbino DR, Diekhans M, et al. 2011. Assemblathon 1: a competitive assessment of de novo short read assembly methods. Genome Res 21: 2224-2241. doi: 10.1101/gr.126599.111

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323: 133-138. doi: 10.1126/science.1162986

Faria PJ, Lazarus CM, van Oosterhout C, Harris PD, Cable J 2011. First polymorphic microsatellites for the gyrodactylids, an important group of fish pathogens. Conservation Genetics Resources 3: 177-180.

Fromm B, Worren MM, Hahn C, Hovig E, Bachmann L 2013. Substantial Loss of Conserved and Gain of Novel MicroRNA Families in Flatworms. Mol Biol Evol 30: 2619-2628. doi: 10.1093/molbev/mst155

Gan HM, Schultz MB, Austin CM 2014. Integrated shotgun sequencing and bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms substantial gene rearrangements in Australian freshwater crayfishes. BMC Evol Biol 14: 19. doi: 10.1186/1471-2148-14-19

Gum B, Gross R, Kuehn R 2005. Mitochondrial and nuclear DNA phylogeography of European grayling (Thymallus thymallus): evidence for secondary contact zones in central Europe. Mol Ecol 14: 1707-1725. doi: 10.1111/j.1365-294X.2005.02520.x

Hansen H, Bachmann L, Bakke TA 2003. Mitochondrial DNA variation of Gyrodactylus spp (Monogenea, Gyrodactylidae) populations infecting Atlantic salmon, grayling, and

rainbow trout in Norway and Sweden. International Journal for Parasitology 33: 1471-1478.

Hansen H, Bakke TA, Bachmann L 2007. Mitochondrial haplotype diversity of Gyrodactylus thymalli (Platyhelminthes; Monogenea): extended geographic sampling in United Kingdom, Poland, and Norway reveals further lineages. Parasitol Res 100: 1389-1394. doi: 10.1007/s00436-006-0423-5

Hansen H, Martinsen L, Bakke TA, Bachmann L 2006. The incongruence of nuclear and mitochondrial DNA variation supports conspecificity of the monogenean parasites Gyrodactylus salaris and G. thymalli. Parasitology 133: 639-650. doi: 10.1017/S0031182006000655

Harris PD 1998. Ecological and genetic evidence for clonal reproduction in Gyrodactylus gasterostei Glaser, 1974. International Journal for Parasitology 28: 1595-1607.

Harris PD 1989. Interactions between population growth and sexual reproduction in the viviparous monogenean Gyrodactylus turnbulli Harris, 1986 from the guppy Poecilia reticulata Peters. Parasitology 98: 245-251.

Harris PD 1993. Interactions between reproduction and population biology in gyrodactylid monogeneans—a review. Francais de la Peche et de la Pisciculture 1: 47-65.

Harris PD 1983. The morphology and life-cycle of the oviparous Oögyrodactylus farlowellae gen. et sp. nov. (Monogenea, Gyrodactylidea). Parasitology 87: 405-420.

Harris PD, Bachmann L, Bakke TA 2011. Freshwater charr (Salvelinus alpinus) as hosts for Gyrodactylus salaris: implications for management. Vet Rec 168: 161. doi: 10.1136/vr.c4744

Harris PD, Jansen PA, Bakke TA 1994. The population age structure and reproductive biology of Gyrodactylus salaris Malmberg (Monogenea). Parasitology 108 ( Pt 2): 167-173.

Harris PD, Shinn AP, Cable J, Bakke TA 2004. Nominal species of the genus Gyrodactylus von Nordmann 1832 (Monogenea: Gyrodactylidae), with a list of principal host species. Systematic Parasitology 59: 1-27. doi: 10.1023/B:SYPA.0000038447.52015.e4

Huyse T, Buchmann K, Littlewood DT 2008. The mitochondrial genome of Gyrodactylus derjavinoides (Platyhelminthes: Monogenea)--a mitogenomic approach for Gyrodactylus species and strain identification. Gene 417: 27-34. doi: 10.1016/j.gene.2008.03.008

Huyse T, Plaisance L, Webster BL, Mo TA, Bakke TA, Bachmann L, Littlewood DT 2007. The mitochondrial genome of Gyrodactylus salaris (Platyhelminthes: Monogenea), a pathogen of Atlantic salmon (Salmo salar). Parasitology 134: 739-747. doi: 10.1017/S0031182006002010

Huyse T, Volckaert FA 2005. Comparing host and parasite phylogenies: gyrodactylus flatworms jumping from goby to goby. Syst Biol 54: 710-718. doi: 10.1080/10635150500221036

Janicki C 1920. Grundinien einer ‘‘Cercomer Theorie’’ zur Morphologie der Trematoden und Cestoden. Festschrift für Zschokke 30: 1-22.

Jansen PA, Bakke TA 1993. Regulatory processes in the monogenean Gyrodactylus salaris Malmberg–Atlantic salmon (Salmo salar L.) association: II. Experimental studies. Fisheries Research 17: 103-114.

Jansen PA, Bakke TA 1991. Temperature-dependent reproduction and survival of Gyrodactylus salaris Malmberg, 1957 (Platyhelminthes: Monogenea) on Atlantic salmon (Salmo salar L.). Parasitology 102 Pt 1: 105-112.

Johnsen BO 1978. The effect of an attack by the parasite Gyrodactylus salaris on the population of salmon parr in the river Lakselva, Misvaer in Northern Norway. Astarte 11: 7-9.

Justine JL 1998. Non-monophyly of the monogeneans? International Journal for Parasitology 28: 1653-1657. doi: 10.1016/S0020-7519(98)00060-5

Kearn GC. 1994. Parasitism and the Platyhelminths. London: Chapman and Hall.

King SD, P. B, Cone DK 2014. Gyrodactylus patersoni n. sp. (Monogenea: Gyrodactylidae) Infecting Atlantic Silverside (Menidia menidia) from Nova Scotia, Canada. Comparative Parasitology 81: 27-32. doi: http://dx.doi.org/10.1654/4647.1

Kleftogiannis D, Kalnis P, Bajic VB 2013. Comparing Memory-Efficient Genome Assemblers on Stand-Alone and Cloud Infrastructures. Plos One. doi: 10.1371/journal.pone.0075505

Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, McVey SD, Radune D, Bergman NH, Phillippy AM 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol 14: R101. doi: 10.1186/gb-2013-14-9-r101

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, et al. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol 30: 693-700. doi: 10.1038/nbt.2280

Koskinen MT, Ranta E, Piironen J, Veselov A, Titov S, Haugen TO, Nilsson J, Carlstein M, Primmer CR 2000. Genetic lineages and postglacial colonization of grayling (Thymallus thymallus, Salmonidae) in Europe, as revealed by mitochondrial DNA analyses. Mol Ecol 9: 1609-1624.

Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M 2013. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Front Genet 4: 237. doi: 10.3389/fgene.2013.00237

Kumar S, Schiffer PH, Blaxter M 2012. 959 Nematode Genomes: a semantic wiki for coordinating sequencing projects. Nucleic Acids Res 40: D1295-1300. doi: 10.1093/nar/gkr826

Kuusela J, Holopainen R, Meinila M, Anttila P, Koski P, Zietara MS, Veselov A, Primmer CR, Lumme J 2009. Clonal Structure of Salmon Parasite Gyrodactylus salaris on a Coevolutionary Gradient on Fennoscandian Salmon (Salmo salar). Annales Zoologici Fennici 46: 21-33.

Kuusela J, Zietara MS, Lumme J 2007. Hybrid origin of Baltic salmon-specific parasite Gyrodactylus salaris: a model for speciation by host switch for hemiclonal organisms. Mol Ecol 16: 5234-5245. doi: 10.1111/j.1365-294X.2007.03562.x

Lautraite A, Blanc G, Thiery R, Daniel P, Vigneulle M 1999. Gyrodactylids parasitizing salmonids in Brittany and Western Pyrénées water basins: Epidemiological features of infection and species composition. Bulletin Français de Pêche et de la Pisciculture 355: 305-325.

Lerceteau-Kohler E, Schliewen U, Kopun T, Weiss S 2013. Genetic variation in brown trout Salmo trutta across the Danube, Rhine, and Elbe headwaters: a failure of the phylogeographic paradigm? BMC Evol Biol 13: 176. doi: 10.1186/1471-2148-13-176

Littlewood DTJ. 2006. The Evolution of Parasitism in Flatworms. In. Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology and Physiology. Cabi Publishing, Wallingford 0x10 8de, Oxon, Uk: Cabi Publishing-C a B Int. p. 1-36.

Littlewood DTJ, Rohde K, Clough KA 1999. The interrelationships of all major groups of Platyhelminthes: phylogenetic evidence from morphology and molecules. Biological Journal of the Linnean Society 66: 75-114. doi: 10.1111/j.1095-8312.1999.tb01918.x

Lockyer AE, Olson PD, Littlewood DTJ 2003. Utility of complete large and small subunit rRNA genes in resolving the phylogeny of the Neodermata (Platyhelminthes): implications and a review of the cercomer theory. Biological Journal of the Linnean Society 78: 155-171.

Malmberg G, Malmberg M 1993. Species of Gyrodactylus (Platyhelminthes, Monogenea) on salmonids in Sweden. Fisheries Research 17: 59-68.

Mariac C, Scarcelli N, Pouzadou J, Barnaud A, Billot C, Faye A, Kougbeadjo A, Maillol V, Martin G, Sabot F, et al. 2014. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. Mol Ecol Resour. doi: 10.1111/1755-0998.12258

Matejusova I, Gelnar M, McBeath AJ, Collins CM, Cunningham CO 2001. Molecular markers for gyrodactylids (Gyrodactylidae: Monogenea) from five fish families (Teleostei). International Journal for Parasitology 31: 738-745.

Meinila M, Kuusela J, Zietara MS, Lumme J 2004. Initial steps of speciation by geographic isolation and host switch in salmonid pathogen Gyrodactylus salaris (Monogenea: Gyrodactylidae). International Journal for Parasitology 34: 515-526. doi: 10.1016/j.ijpara.2003.12.002

Metzker ML 2010. Sequencing technologies - the next generation. Nature Reviews Genetics 11: 31-46. doi: 10.1038/nrg2626

Miller JR, Koren S, Sutton G 2010. Assembly algorithms for next-generation sequencing data. Genomics 95: 315-327. doi: 10.1016/j.ygeno.2010.03.001

Nagarajan N, Pop M 2013. Sequence assembly demystified. Nature Reviews Genetics 14: 157-167. doi: 10.1038/nrg3367

Nosil P 2002. Transition rates between specialization and generalization in phytophagous insects. Evolution 56: 1701-1706.

Olson PD, Tkach VV 2005. Advances and trends in the molecular systematics of the parasitic platyhelminthes. Advances in Parasitology, Vol 60 60: 165-243. doi: 10.1016/S0065-308x(05)60003-6

Olstad K, Bachmann L, Bakke TA 2009. Phenotypic plasticity of taxonomic and diagnostic structures in gyrodactylosis-causing flatworms (Monogenea, Platyhelminthes). Parasitology 136: 1305-1315. doi: 10.1017/S0031182009990680

Olstad K, Cable J, Robertsen G, Bakke TA 2006. Unpredicted transmission strategy of Gyrodactylus salaris (Monogenea: Gyrodactylidae): survival and infectivity of parasites on dead hosts. Parasitology 133: 33-41. doi: 10.1017/S0031182006009966

Olstad K, Robertsen G, Bachmann L, Bakke TA 2007a. Variation in host preference within Gyrodactylus salaris (Monogenea): an experimental approach. Parasitology 134: 589-597. doi: 10.1017/S0031182006001715

Olstad K, Shinn AP, Bachmann L, Bakke TA 2007b. Host-based identification is not supported by morphometrics in natural populations of Gyrodactylus salaris and G. thymalli (Platyhelminthes, Monogenea). Parasitology 134: 2041-2052. doi: 10.1017/S0031182007003332

Paladini G, Gustinelli A, Fioravanti ML, Hansen H, Shinn AP 2009. The first report of Gyrodactylus salaris Malmberg, 1957 (Platyhelminthes, Monogenea) on Italian cultured stocks of rainbow trout (Oncorhynchus mykiss Walbaum). Vet Parasitol 165: 290-297. doi: 10.1016/j.vetpar.2009.07.025

Park JK, Kim KH, Kang S, Kim W, Eom KS, Littlewood DTJ 2007. A common origin of complex life cycles in parasitic flatworms: evidence from the complete mitochondrial genome of Microcotyle sebastis (Monogenea : Platyhelminthes). BMC Evol Biol 7. doi: 10.1186/1471-2148-7-11

Pell J, Hintze A, Canino-Koning R, Howe A, Tiedje JM, Brown CT 2012. Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. Proc Natl Acad Sci U S A 109: 13272-13277. doi: 10.1073/pnas.1121464109

Perkins EM, Donnellan SC, Bertozzi T, Whittington ID 2010. Closing the mitochondrial circle on paraphyly of the Monogenea (Platyhelminthes) infers evolution in the diet of parasitic flatworms. International Journal for Parasitology 40: 1237-1245. doi: 10.1016/j.ijpara.2010.02.017

Plaisance L, Huyse T, Littlewood DT, Bakke TA, Bachmann L 2007. The complete mitochondrial DNA sequence of the monogenean Gyrodactylus thymalli (Platyhelminthes: Monogenea), a parasite of grayling (Thymallus thymallus). Mol Biochem Parasitol 154: 190-194. doi: 10.1016/j.molbiopara.2007.04.012

Ramirez R, Bakke TA, Harris PD 2014. Same barcode, different biology: differential patterns of infectivity, specificity and pathogenicity in two almost identical parasite strains. International Journal for Parasitology in press.

Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, et al. 2011. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med 365: 709-717. doi: 10.1056/NEJMoa1106920

Robertsen G, Hansen H, Bachmann L, Bakke TA 2007. Arctic charr (Salvelinus alpinus) is a suitable host for Gyrodactylus salaris (Monogenea, Gyrodactylidae) in Norway. Parasitology 134: 257-267. doi: 10.1017/S0031182006001223

Rokicka M, Lumme J, Zietara MS 2007. Identification of Gyrodactylus ectoparasites in Polish salmonid farms by PCR-RFLP of the nuclear ITS segment of ribosomal DNA (Monogenea, Gyrodactylidae). Acta Parasitologica 52: 185-195.

Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, Treangen TJ, Schatz MC, Delcher AL, Roberts M, et al. 2012. GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Res 22: 557-567. doi: 10.1101/gr.131383.111

Schelkle B, Faria PJ, Johnson MB, van Oosterhout C, Cable J 2012. Mixed infections and hybridisation in monogenean parasites. Plos One 7: e39506. doi: 10.1371/journal.pone.0039506

Shinn AP, Collins C, Garcia-Vasquez A, Snow M, Matejusova I, Paladini G, Longshaw M, Lindenstrom T, Stone DM, Turnbull JF, et al. 2010. Multi-centre testing and validation of current protocols for the identification of Gyrodactylus salaris (Monogenea). International Journal for Parasitology 40: 1455-1467. doi: 10.1016/j.ijpara.2010.04.016

Shinn AP, Hansen H, Olstad K, Bachmann L, Bakke TA 2004. The use of morphometric characters to discriminate specimens of laboratory-reared and wild populations of Gyrodactylus salaris and G. thymalli (Monogenea). Folia Parasitol (Praha) 51: 239-252.

Soleng A, Bakke TA 2001. The susceptibility of grayling (Thymallus thymallus) to experimental infections with the monogenean Gyrodactylus salaris. International Journal for Parasitology 31: 793-797.

Sterud E, Mo TA, Collins CM, Cunningham CO 2002. The use of host specificity, pathogenicity, and molecular markers to differentiate between Gyrodactylus salaris Malmberg, 1957 and G. thymalli Zitnan, 1960 (Monogenea: Gyrodactylidae). Parasitology 124: 203-213.

Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, Tracey A, Bobes RJ, Fragoso G, Sciutto E, et al. 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496: 57-63. doi: 10.1038/Nature12031

van der Werf MJ, de Vlas SJ, Brooker S, Looman CWN, Nagelkerke NJD, Habbema JDF, Engels D 2003. Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. Acta Trop 86: 125-139. doi: 10.1016/S0001-706x(03)00029-9

Vignon M 2011. Putting in shape: towards a unified approach for the taxonomic description of monogenean haptoral hard parts. Systematic Parasitology 79: 161-174. doi: 10.1007/s11230-011-9303-1

von Gersdorff Jorgensen L, Heinecke RD, Kania P, Buchmann K 2008. Occurrence of gyrodactylids on wild Atlantic salmon, Salmo salar L., in Danish rivers. J Fish Dis 31: 127-134. doi: 10.1111/j.1365-2761.2007.00875.x

Weiss S, Persat H, Eppe R, Schlotterer C, Uiblein F 2002. Complex patterns of colonization and refugia revealed for European grayling Thymallus thymallus, based on complete sequencing of the mitochondrial DNA control region. Mol Ecol 11: 1393-1407.

Weiss S, Schlotterer C, Waidbacher H, Jungwirth M 2001. Haplotype (mtDNA) diversity of brown trout Salmo trutta in tributaries of the Austrian Danube: massive introgression of Atlantic basin fish--by man or nature? Mol Ecol 10: 1241-1246.

Wu CI, Ting CT 2004. Genes and speciation. Nature Reviews Genetics 5: 114-122. doi: 10.1038/nrg1269

Yandell M, Ence D 2012. A beginner's guide to eukaryotic genome annotation. Nature Reviews Genetics 13: 329-342. doi: 10.1038/nrg3174

Zheng HJ, Zhang WB, Zhang L, Zhang ZZ, Li J, Lu G, Zhu YQ, Wang YZ, Huang Y, Liu J, et al. 2013. The genome of the hydatid tapeworm Echinococcus granulosus. Nat Genet 45: 1168-U1302. doi: 10.1038/Ng.2757

Zietara MS, Kuusela J, Lumme J 2006. Escape from an evolutionary dead end: a triploid clone of Gyrodactylus salaris is able to revert to sex and switch host (Platyhelminthes, Monogenea, Gyrodactylidae). Hereditas 143: 84-90. doi: 10.1111/j.2006.0018-0661.01956.x

Zietara MS, Lumme J 2002. Speciation by host switch and adaptive radiation in a fish parasite genus Gyrodactylus (Monogenea, Gyrodactylidae). Evolution 56: 2445-2458.

# INDIVIDUAL PAPERS