SPECIAL REVIEW

Integrating molecular tools into freshwater ecology: developments and opportunities

STEFFEN U. PAULS*, †, 1, MARIA ALP‡, §, MIKLÓS BÁLINT*, †, PAOLA BERNABÒ‡, **, FEDOR ČIAMPOR JR††, ZUZANA ČIAMPOROVÁ-ZATÓVIČOVÁ††, DEBRA S. FINN‡‡, JAN KOHOUT††, §§, FLORIAN LEESER¶¶, VALERIA LENCIONI§, IVAN PAZ-VINAS§, ***, †††, AND MICHAEL T. MONAGHAN‡, 1

* Biodiversity and Climate Research Centre (BiK-F), Frankfurt am Main, Germany
† Senckenberg Gesellschaft für Naturforschung, Frankfurt am Main, Germany
‡ Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany
§ UMR5174 EDB (Laboratoire Evolution & Diversité Biologique), Ecole Nationale de Formation Agronomique (ENFA), Centre National de la Recherche Scientifique (CNRS), Universite Paul Sabatier, Toulouse Cedex 4, France
¶ Section of Invertebrate Zoology and Hydrobiology, Museo delle Scienze, Trento, Italy
** Centre for Integrative Biology, University of Trento, Trento, Italy
†† Institute of Zoology, Slovak Academy of Sciences, Bratislava, Slovakia
‡‡ School of Geography, Earth & Environmental Sciences, University of Birmingham, Edgbaston, U.K.
§§ Faculty of Fisheries and Protection of Waters, University of South Bohemia in České Budějovice, Vodňany, Czech Republic
¶¶ Department of Animal Ecology, Evolution and Biodiversity, Ruhr University Bochum, Bochum, Germany
*** Centre National de la Recherche Scientifique (CNRS), Station d’Ecologie Expérimentale du CNRS à Moulis, USR2936 Moulis, France
††† Université de Toulouse, UPS, UMR5174 (EDB), Toulouse Cedex 4, France

SUMMARY

1. Molecular genetic techniques have been used in freshwater biology for more than 30 years. Early work focussed on studies of population structure, systematics and taxonomy. More recently, the range of studies has broadened to include ecology and adaptation. Advances in analytical methods and in technology (e.g. next-generation sequencing) and decreasing costs of data production ensure that the field will continue to develop and broaden in scope.

2. At least three factors make the application of molecular techniques to freshwater biology exciting. First, the highly variable nature of many aquatic habitats makes them excellent models for the study of environmental change on ecological and evolutionary time scales. Second, the mature state of the field of freshwater biology provides an extensive foundation of ecological knowledge of freshwater organisms and their distinct adaptations. Third, the methodological advances allow researchers to focus more on merging molecular and ecological research and less on designing studies around technical limitations.

3. We identified eight research areas in freshwater biology in which the integration of molecular and ecological approaches provides exceptional opportunities. The list is not exhaustive, but considers a broad range of topics and spans the continuum from basic to applied research. The areas identified use a combination of natural, experimental and in silico approaches.

4. With advancing molecular techniques, freshwater biology is in an unusually strong position to link the genetic basis and ecological importance of adaptations across a wide range of taxa, ecosystems and spatiotemporal scales. Our aim was to identify opportunities for the integration of molecular and ecological approaches, to motivate greater collaboration and crossover, and to promote exploitation of the synergies of bridging ecological and evolutionary freshwater research.

Correspondence: Steffen Pauls, Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany. E-mail: steffen.pauls@senckenberg.de

1 These authors contributed equally to this work.
2 Present address: Department of Zoology, Oregon State University, Corvallis, OR 97331, U.S.A.
3 [Correction added on 14 May 2014, after first online publication: complete address of affiliation § is now provided]
**Introduction**

The use of molecular techniques to quantify genetic variation at the level of DNA sequences has revolutionised most fields of biology, including systematics, physiology, biochemistry, evolutionary biology and ecology. Recent advances in these techniques have the potential to more closely integrate ecological and evolutionary research (Andrew et al., 2013). Next-generation sequencing (NGS, for further information on terms see Table 1) is perhaps the most promising advance, facilitating the analysis of large numbers of samples, the study of poorly preserved and old samples (e.g. ancient DNA; Knapp & Hofreiter, 2010) and the screening of unprecedented numbers of

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Definition of some technical terms used in this review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate Bayesian Computation (ABC)</td>
<td>ABC algorithms are simulation-based procedures that compare summary statistics (e.g. heterozygosity, allelic richness) calculated from a large number of simulated gene genealogies generated under a given model with those observed in empirical data for inferential and model selection purposes (Csillery et al., 2010)</td>
</tr>
<tr>
<td>Candidate loci</td>
<td>Areas of the genome with known or inferred function and often hypothesized to be causal contributors to genetic variation in a quantitative trait (Hamilton, 2009)</td>
</tr>
<tr>
<td>Census population size (Nc)</td>
<td>The number of individuals in a population; the head-count size of a population (Hamilton, 2009)</td>
</tr>
<tr>
<td>Coalescent</td>
<td>A retrospective model of population genetics, where a group of lineages in the present is connected back through time to a single ancestor (MRCA = most recent common ancestor); the connection of lineages in the past is a coalescent event (Hamilton, 2009)</td>
</tr>
<tr>
<td>DNA barcoding</td>
<td>A technique for identifying species using a short DNA sequence from a standardized locus (e.g. COI in animals; Hebert et al., 2003)</td>
</tr>
<tr>
<td>Effective population size (Ne)</td>
<td>The number of breeding individuals in an idealized population that would show the same amount of dispersion of allele frequencies under random genetic drift or the same amount of inbreeding as the population under consideration (Wright, 1938). Ne is usually lower than Nc and is driven by successfully breeding individuals of a population</td>
</tr>
<tr>
<td>Environmental DNA</td>
<td>A mixture of genomic DNA from different organisms that is extracted directly from environmental samples, e.g. water, and not from selected target organisms (Taberlet et al., 2012)</td>
</tr>
<tr>
<td>Genome annotation</td>
<td>The process during which biological information on gene structure and function are added to specific genomic regions. This is done by comparing the similarity of the gene in question to genes with a known function; function may also be inferred from gene expression data under specific environmental conditions</td>
</tr>
<tr>
<td>Genomics</td>
<td>Used here to refer to all genetic approaches that explicitly investigate the structure and function of the whole genome, rather than methods that target individual loci</td>
</tr>
<tr>
<td>Individual-based forward simulators</td>
<td>Algorithms that generate genetic data for simulated individuals in progressively forward-in-time generations. The genetic state at a generation T + 1 is determined by: (i) the state at the generation T (i.e. the current generation) and (ii) by a succession of transition matrices that characterize the life cycle of individuals (e.g. birth, reproduction, migration, death; see Hoban et al., 2012). Better suited for predictive analyses</td>
</tr>
<tr>
<td>Next-generation sequencing (NGS)</td>
<td>High-throughput sequencing technologies that parallelise the sequencing process by spatially separating clonally amplified DNA templates or single DNA molecules; generally produce millions of sequences concurrently</td>
</tr>
<tr>
<td>Model/Non-model organism</td>
<td>Model organisms have served as pioneers for genomics and most genomic resources were first developed for model organisms. They were chosen based on their simple life histories and ease with which they are reared in the laboratory. In contrast, here we refer to species that live in wild, natural populations and that are captured from these for experiments and genetic or genomic study as non-model organisms</td>
</tr>
<tr>
<td>Population-based coalescent simulators</td>
<td>Algorithms that generate genetic data for populations backwards in time. Gene lineages are traced back in time to the most recent common ancestor. The coalescence of lineages is influenced by the characteristics of the model (e.g. historical events such as bottlenecks, migration rates). Better suited for inferential questions at an evolutionary timescale (Hoban et al., 2012)</td>
</tr>
<tr>
<td>Transcriptome</td>
<td>The set of all RNA molecules in an organism, tissue or cell, representing those parts of the genome that are actively being transcribed into RNA molecules or ‘expressed’ at a given point in time</td>
</tr>
</tbody>
</table>
genetic markers (Perry & Rowe, 2011). One of the most important results from the growth of NGS is the large-scale collection of data from non-model organisms (Table 1). Many large-scale initiatives are generating comparable data across taxa and ecosystems that were, until recently, only available for a few model groups (e.g. International Barcode of Life: ibol.org; Insect and other Arthropod Genome Sequencing Initiative: arthropodgenomes.org/wiki/i5K; 1K Insect Transcriptome Evolution: 1kite.org, Microbial Genome Database: mbgd.genome.ad.jp).

Freshwater biologists have used molecular tools for more than 40 years to address questions about population structure and evolutionary history, for taxonomic and systematics studies, and more recently for biomonitoring. Our search of the literature (Fig. 1) shows a steady increase in the use of molecular data in freshwater research since about 1980. Nearly all of these studies have used putatively neutral markers to survey genetic variation in natural populations (Hughes, Schmidt & Finn, 2009), with little direct integration into ecological experiments. The increasing number of markers and improved ability to study functional, potentially adaptive genetic variation frees scientists from earlier methodological constraints (Andrew et al., 2013). The result is the potential to focus more on ecological questions and less on designing studies based on technical limitations.

In their review, Andrew et al. (2013) largely focussed on terrestrial species and ecosystems, with very few freshwater examples. Nonetheless, a number of characteristics of freshwater species and ecosystems make them fascinating systems for ecological research that integrates molecular tools. Fresh waters cover only 1% of the Earth’s surface but are home to 6% of all insect species, the most evolutionarily successful and diverse group of animal life (Dijkstra, Monaghan & Pauls, 2014). Little is known about the processes that led to this remarkable diversification; however, the extensive ecological research carried out on many freshwater groups provides a strong basis for hypothesis-driven research into the mechanisms that generate and maintain freshwater diversity (Dijkstra et al., 2014). The physical environment of fresh waters requires adaptations to breathing under water and to an energy base often composed mainly of allochthonous carbon inputs, because light for primary production is limited in deep or murky water and because of the extensive land–water interface in smaller freshwater bodies and river systems compared with the oceans. By applying new molecular tools (both technological and analytical) to freshwater science, we can (i) characterise spatial patterns of genetic diversity in fresh waters with greater resolution in a broader range of taxa; (ii) characterise functional genetic variation and assess levels of adaptation and responses to environmental changes; (iii) assess the current and project the future status of freshwater biodiversity with unprecedented speed, detail and taxonomic resolution; and (iv) integrate this new information to improve freshwater management and conservation strategies.

The aim of this paper is to highlight the potential of molecular tools in supporting the study of some key topics in freshwater ecology. We identify eight concepts that span questions and applications and can be applied across the continuum of basic and applied research (Fig. 2). Ours is not an exhaustive review and does not cover some other emerging or important topics in freshwater molecular ecology, such as microbial diversity (Pernthaler, 2013) or ecological diversification in fish.

![Fig. 1 Results from a Web of Science](Thomson Reuters, 2014) search using the keywords Topic = molecule* OR genetic* AND Topic = freshwater OR fresh water OR fresh-water resulted in 10 032 records on 12 February 2014. Shown are absolute article record counts (bars) and the proportion of freshwater papers with molecular as a key word (open circles and line).
Eight topics for integrating molecular tools in freshwater ecology

Gene–environment interactions

One of the ways in which organisms respond to short-term changes in their environment is by varying the level of expression of a given gene or genes in the cell (López-Maury, Marguerat & Bähler, 2008). Changes in gene expression are determined from differences in the number of messenger RNA (mRNA) molecules for a given gene that are found in a tissue. Studying how and when these changes occur is an important component of developmental biology, molecular physiology and ecotoxicology and may improve our understanding of how environmental changes affect natural populations (Kultz, 2005; Travers et al., 2007; Hoffmann & Willi, 2008). When combined with an appropriate experimental setup (see topic ‘Integration of experimental and genetic approaches’), studies of gene expression can help distinguish adaptive responses to environmental change from phenotypically plastic responses (Haap & Köhler, 2009; Scoville & Pfrender, 2010; Latta et al., 2012). This latter point is particularly important because we know little about phenotypic plasticity in response to stress in non-model organisms. In fact, most of our knowledge of freshwater invertebrates (Schmidt-Kloiber & Hering, 2012) is based on correlations of species occurrence and not on direct measures of tolerance or preference (Pauls et al., 2013).

To date, most analyses of gene expression have been based on changes in expression of one or more candidate genes (Table 1) that are thought to play a role in responses to environmental change. Changes in expression could reflect instantaneous response, acclimation over hours to months or adaptation over generations (Schulte, 2004). For instance, increased expression of genes involved in the synthesis of heat-shock proteins is typically used as an indicator of environmental stress. These proteins are often more highly expressed (‘up-regulated’) in aquatic organisms during times of environmental stress such as temperature extremes and exposure to toxins. A series of studies of thermal tolerance in alpine Chironomidae examined expression of the hsp70 gene in larvae under natural conditions (Lencioni, Boschini & Rebecchi, 2009) and under controlled stressed conditions (Bernabò et al., 2011; Lencioni et al., 2013). Using both short-term heat shocks and longer-term high-temperature exposures, these studies found pronounced interspecific and interseasonal variation in hsp70 expression and, in one of the species, a positive correlation with survival rate under short-term heat shock. These results shed light on the regulation of physiological responses to stress and provide new insights into possible responses of freshwater insects to global warming.

One advance resulting from the wider availability of genomic data is an increase in the scope of expression studies particularly regarding the number of candidate loci that can be tested. Recent work on Daphnia pulex highlights both aspects of these advances. D. pulex (Santos & Salzburger, 2012). For the eight topics identified, we briefly summarise work that has been done and outline further developments related to molecular techniques that, in our opinion, may considerably enrich freshwater science. The authors’ experience and focus mean that many of the examples are taken from stream ecology, but most of the concepts are applicable to any type of freshwater ecosystem. We hope to stimulate discussion and interaction among freshwater scientists in diverse fields, using a range of methodologies. The ultimate aims are to devise strategies for jointly tackling questions of interest for freshwater and molecular ecologists, and to generate new syntheses.
displays different phenotypes in varying environments, including morphological defence structures in the presence of predation threat (Krueger & Dodson, 1981; Petrusek et al., 2009). Genetically identical clones can differ substantially under different environmental conditions, and the mechanisms underlying this plasticity have puzzled ecologists and evolutionary biologists for decades (Via et al., 1995). Miyakawa et al. (2010) screened the D. pulex genome available on public databases to identify 31 candidate loci for a subsequent study of genes expressed during D. pulex development. They found several genes from different biosynthetic pathways that were more highly expressed during D. pulex development in the presence of predator cues.

A second advance from NGS capabilities is the ability to sequence the entire genome of an organism. In addition with NGS or microarray methods the transcriptome, that is, all expressed genes can be profiled. This allows researchers to screen thousands of genes for increased (or decreased) expression under certain conditions and potentially identify new genes in a genome that are important. Colbourne et al. (2011) applied NGS methods to sequence the genome of D. pulex to characterise signatures of adaptation in different regions of the genome. In addition, they applied a whole-genome microarray approach to investigate gene expression levels under different stress conditions. A large proportion of the genes that were differently expressed in D. pulex exposed to stressors were found in genomic regions for which there was no previous knowledge of gene identity and function (Colbourne et al., 2011; Latta et al., 2012). A major proportion of this ‘eco-responsive genome’ would therefore have gone unnoticed when relying on known candidate loci alone, even when studying multiple candidate loci. Another central finding of sequencing the genome was that a great number of differentially-regulated Daphnia-specific gene families have emerged recently via recent gene duplication (Colbourne et al., 2011). These duplicated genes presumably play a central functional role by enabling a specific and fine-tuned production of gene products in response to different environmental conditions and may, in part, explain the ecological and evolutionary success of Daphnia.

The approaches described above come under the general umbrella of ecological genomics. Ecological genomics can be broadly defined as the study of: (i) interactions of many genes with the environment; (ii) structural changes in the genome in the context of adaptation processes or (iii) genome-wide scans for ecologically relevant genes without prior knowledge. This field thus seeks to link organismal short-term responses and long-term adaptation to environmental cues with genomic patterns (e.g. gene structure) and processes (e.g. transcriptional activity) by extending studies from model organisms (Table 1) to natural populations (Ungerer, Johnson & Herman, 2008). Most of our current understanding of genome-environment interactions stems from laboratory model organisms (e.g. Danio spp., Drosophila spp., Caenorhabditis elegans). Many studies of these species give a poor representation of natural systems because of inbreeding depression in laboratory populations and experimental conditions that do not always reflect natural habitats (Peña-Castillo & Hughes, 2007). An important benefit of NGS and related technologies (Gardner et al., 2011) is that organisms for genomic research can now be selected for their interesting ecology or evolutionary history and not only due to their suitability for laboratory studies. Consequently, recent whole-genome sequences were generated for interesting freshwater species including the aforementioned D. pulex (Colbourne et al., 2011) and the three-spined stickleback Gasterosteus aculeatus (Jones et al., 2012). For these and other species in the future, freshwater ecologists will now be able to identify the combinations and interactions of the genes involved in responses to environmental conditions, as well as those genes that enabled adaptation to the special habitat conditions in fresh waters, like osmotic regulation in fish and benthic invertebrates or changes in thermal acclimation for insects with aquatic larvae and terrestrial adults. Investigation of quantitative relationships between changes in an environmental parameter and gene expression can then focus on applying quantitative PCR (qPCR) or microarrays of selected candidate loci from the set of responsive genes (e.g. Murdoch, Moller-Jacobs & Thomas, 2013).

The increasing amount of genomic data available leads to new challenges, perhaps most importantly in the computational demands of analysing such large data sets (Carstens, Lemmon & Lemmon, 2012) and the assignment of biological function to sequence data, that is, functional gene annotation (Table 1). Most genes are assigned a function based on their similarity to other genes whose function has already been described (usually in Drosophila melanogaster or Mus musculus); however, without validating the function of those genes experimentally, we run the risk of inferring the wrong function. This is particularly true for the large proportion of regulatory genome regions (Jones et al., 2012). Experimental approaches that connect observed genomic patterns with biological functions are therefore of central importance (see topic ‘Integration of experimental and genetic approaches’).
**Integration of experimental and genetic approaches**

Molecular techniques can deliver valuable ecological and evolutionary information; however, most studies are inherently descriptive, and understanding causalities is not always straightforward. This is likely to be a consequence of many molecular ecology studies in freshwater systems being based on surveying genetic data in space, and sometimes in time. For example, a large body of literature has resulted from DNA barcoding (see topic ‘Community ecology using DNA-metabarcoding’), and from phylogeographic (see topic ‘Comparative population structure’) and population genetic studies (see topics ‘Linking population genetics and community ecology’ and ‘Population genetics of invasive and managed species’). These studies can reveal important and interesting patterns on the distribution of genetic diversity in space and time, but generally do not fully exploit the potential of molecular data to reveal underlying processes. One reason for this is that a large number of environmental and demographic factors affect the distribution of genetic diversity. Laboratory or field experiments can help researchers focus on a manageable set of parameters by which to explain observed patterns.

For example, in a series of multigeneration experiments on the non-biting midge *Chironomus riparius*, genetic (microsatellite) diversity was analysed as a response variable to multiple environmental stressors (temperature and toxins). In one experiment, regional populations between Portugal and Germany were reared under different temperature conditions, revealing that growth rates were linked not only to thermal conditions but also to the genetic diversity of the populations (Nemec et al., 2013). Other experiments showed that neutral genetic diversity was reduced over generations under exposure to certain toxins (Nowak et al., 2009; Vogt et al., 2010) and that genetic diversity was most strongly affected when populations were exposed to a combination of low toxin concentrations and thermal stress than to either stressor individually (Müller et al., 2012). When inbred and genetically diverse populations were exposed to different thermal conditions, thermal stress led to reduced fitness in females (reduced number of eggs per egg mass) of the inbred, but not in the genetically diverse populations (C. Nowak, unpubl. data). Although this latter experiment did not directly assess functional diversity related to fitness parameters, it linked a reduction in genetic diversity to negative fitness effects under stress conditions.

Another example highlighting the importance and benefit of marrying molecular and experimental studies is the evolutionary ecology of invasion of the salt water copepod *Eurytemora affinis* into fresh water. Here, the identification of distinct mitochondrial lineages (one invasive and one not) in the St. Lawrence River Seaway (Winkler, Dodson & Lee, 2008) gave rise to a series of experiments aimed at understanding the mechanisms that allow some specific populations/lineages to invade fresh waters. An experiment showed that high food concentration in fresh waters may have allowed a saline ancestor to colonise fresh waters, upon which this invading lineage subsequently evolved an increased tolerance of low salinity, even where food was scarce (Lee et al., 2013). Furthermore, common garden experiments integrating gene expression analysis (see topic ‘Gene-environment interactions’) linked the adaptation to changes in ion uptake and differential ATPase activity (Lee et al., 2011). In cases like this, the combination of molecular analyses with experimental approaches can substantially improve studies by allowing firmer conclusions about the causes of the patterns found.

**Comparative population structure**

Population genetic and phylogeographic studies examine the spatial (and sometimes temporal) distribution of genetic variants (e.g. alleles or genotypes). These genes are neutral (i.e. not under selection), and their spatial patterns are used to address topics such as gene flow, evolutionary history and biogeography (Smith, McVeagh & Collier, 2006; Brändle et al., 2007; Braaker & Heckel, 2009; Kubow et al., 2010; Čiamporová-Zalovičová & Čiampor, 2011) or species’ dispersal (Engelhardt, Haase & Pauls, 2011; Hughes et al., 2011; Yaegashi et al., 2014). While most studies have focussed on one or a few species and one or a few genes, an emerging concept is the shift away from the study of single species to the study of many species simultaneously in a comparative framework to extract general patterns (Andrew et al., 2013). This shift is driven largely by NGS and greater data availability and applicability to non-model organisms, as well as analytical developments (see topic ‘Statistical simulation for assessing population structure’).

Carefully selecting the taxa for such a comparative population structure assessment can reveal the relative importance of particular biological traits on species and population history (Heilveil & Berlocher, 2006; Hodges, Rowell & Keogh, 2007). For example, comparing species with differing dispersal-related traits (life history, dispersal capacity) but similar present-day distributions can reveal the importance of these traits on the historical and present patterns of gene flow among populations (Hughes, 2007; Zickovich & Bohonak, 2007; Lehrian,
Pauls & Haase, 2009; Alp et al., 2012). Assessing the population genetic structure of codistributed taxa with phylogeographic methods can also reveal general biogeographic patterns for regional biotas. Hughes et al. (2011), for example, showed that three species of aquatic insects experienced a common population divergence event at the same time in the Canondale ranges of Australia and that this event may have been triggered by Pleistocene climate dynamics. In contrast, Theissinger et al. (2011, 2013) showed that two European arctic-alpine aquatic insect species with similar present-day distribution patterns experienced different Pleistocene histories and that the post-glacial colonisation of the Scandinavian populations originated in very different refugial areas.

While there is a lot of potential in comparative population genetic and phylogeographic approaches, most studies to date show that patterns are generally species-specific (Stewart et al., 2010), at least at the scale of comparisons that are usually used (<10 species). Generating much larger data sets for greater numbers of species will be important if we wish to reveal general or repeated patterns. Several studies have achieved this for aquatic insects (Isambert et al., 2011; Bergsten et al., 2012; Baseïga et al., 2013), but most studies are based on a single locus (e.g. mitochondrial DNA) and are thereby limited in the conclusions that can be made (but see Monaghan et al., 2009). Multilocus studies provide insight into the evolutionary patterns of species, rather than single genes (Nichols, 2001), are less affected by clonal inheritance, yet still allow the assessment of haplotype ratios to infer gene flow and population genetic structure (Schultheis et al., 2014). They can thus provide more reliable insights into the processes underlying present-day population genetic structure (Elbrecht et al., 2014). Given the current technological capabilities for identifying variable nuclear loci (Schultheis et al., 2014) and the development in multilocus analyses (see topic ‘Statistical simulation for assessing population structure’), future work on comparative population structure should be based on multiple loci when possible.

Linking population genetics and community ecology

Freshwater ecology has a long history of describing spatial patterns of biological diversity (Allan & Castillo, 2010), and molecular tools have provided a great deal of information on genetic diversity (Gustafson et al., 2007; Bálint et al., 2011, 2012) and population structure of freshwater species (Bunn & Hughes, 1997; Monaghan et al., 2002; Wilcock, Nichols & Hildrew, 2003). The study of genetic and species diversity in fresh waters has often been carried out separately, despite several conceptual and methodological similarities (Etienne & Olff, 2004; Hu, He & Hubbell, 2006). Many of the research questions are analogous, quantifying spatial patterns of diversity in order to assess the effects of isolation, history and human impact. Recognition of these conceptual parallels has led to a variety of hypotheses that typically predict a correlation between genetic and species richness and suggest they are influenced by similar processes (Vellend, 2005; Finn & Poff, 2011; Papadopoulou et al., 2011; Schultheis et al., 2012; Múrria et al., 2013).

Stream habitats are spatially organised in a hierarchical, dendritic structure. They are subject to unidirectional gradients such as slope and water flow, but otherwise very stochastic physical dynamics. Longitudinal patterns of genetic and community diversity might respond in a parallel manner (Finn & Poff, 2011), and some recent studies have begun to bridge the gap between approaches at community and population genetic levels to test this hypothesis. Sei, Lang & Berg (2009) and Bonada et al. (2009) measured taxonomic diversity of a series of communities concurrently with genetic diversity of one or two common species from those communities. Sei et al. (2009) reported that geographical distance between habitats was related to genetic distances between fish (Gambusia nobilis) and amphipod (Gammarus spp.) populations, as well as to similarity of the macroinvertebrate communities. They concluded that spatial isolation was the determining factor at both genetic and species levels. Bonada et al. (2009) assessed community diversity patterns of Trichoptera and population genetic patterns of a common species (Chimarra marginata) in the western Mediterranean basin. They found that community and population genetic patterns (of C. marginata) were only related in smaller streams and concluded that headwater communities were more likely to retain the historic population–community relationship, with larger streams having become more homogenised across the western Mediterranean. It is becoming feasible not only to compare community assessments with genetic structure of single or a few species, but also to assess the genetic structure of entire assemblages. For example, Múrria et al. (2013) also found homogenisation in mid-river reaches and greater distinction among headwaters by assessing the population genetic structure of the 17 species of Hydropsyche in the eastern Iberian Peninsula.

Although the fundamental units are analogous (taxa in communities and assemblages, alleles in populations), different approaches to quantifying their diversity have made comparisons unwieldy. Genetic differences among
populations are most often reported as a fixation index ($F_{ST}$) which quantifies the proportion of total genetic variation that results from differences among populations. In contrast, taxonomic differences among communities are often reported using dissimilarity statistics (e.g. Jaccard, Bray–Curtis or Sørensen indices) that do not incorporate the contribution of within-site variation (Finn et al., 2011). Two recent studies applied fixation indices to both species and genetic turnover for a more direct comparison. Finn & Poff (2011) revealed significant substructure ($F_{ST}$) at both levels in high-alpine headwater streams, suggesting that physical isolation acts at both biological levels across a relatively small spatial extent. Evanno et al. (2009) showed that $F_{ST}$ for both freshwater gastropod assemblages and for populations of one common species increased (i.e. were reduced in similarity) following a drought disturbance in a large French floodplain. The authors concluded that drought affected the structure of communities and populations similarly, through a combination of neutral and selective processes. Finn et al. (2011) compared diversity patterns at both community and population genetic levels between ‘headwaters’ (stream orders one-two) and ‘mid-order’ reaches (orders three-four) within ecoregions. They found that headwaters were more highly structured (more different from one another) than mid-order streams at this spatial scale, both in terms of community and population genetic diversity. Additionally, dissimilarity indices gave a stronger signal of these patterns than fixation, which led the authors to conclude that the traditional community-level dissimilarity analyses might be more appropriate than $F_{ST}$ when strictly studying differentiation among sites with genetic diversity data.

It is becoming evident that the fields of population genetics and community ecology for assessing the spatial distribution of biodiversity should probably not continue as independent lines of inquiry. The merging of these traditionally isolated fields has only just begun to shed light on the extent to which species and genetic diversity are influenced by similar processes in time and space (Bonada et al., 2009; Sei et al., 2009; Finn et al., 2011; Baselga et al., 2013). Now is a good time to pursue this. Previous population genetic studies were often limited to examining few molecular loci. With NGS technologies, it is now possible to assess the diversity and distribution of 100s or 1000s of independent genetic loci. This represents a paradigm shift from the study of many individuals at a few loci to examining a few individuals at many loci (Andrew et al., 2013). As fresh waters face increasing anthropogenic pressure (Strayer & Dudgeon, 2010), it is critical that ecologists develop effective ways for understanding and predicting how habitat and environmental change will influence freshwater biodiversity simultaneously at the level of genes and species.

**Statistical simulation for assessing population structure**

Many processes in evolution and ecology are too complex to assess with manageable experiments or surveys. In such cases, simulation-based modelling, that is, the generation of simulated data sets, can offer an alternative approach. In population genetics, simulation-based modelling refers to the generation of simulated genetic datasets (i.e. genotypic data) under pre-defined evolutionary and/or demographic models, mainly for predictive or inferential purposes (Hoban, Bertorelle & Gaggiotti, 2012). Simulation-based modelling is increasingly important in population genetics because it helps to isolate the role of individual processes in determining complex patterns of evolution (Balkenhol, Waits & Deziani, 2009; Balkenhol & Landguth, 2011; Hoban et al., 2012). In freshwater ecology, simulation-based modelling can help us to understand and predict genetic diversity patterns in spatially complex ecosystems such as dendritic stream networks or isolated lakes. They are thus particularly valuable in landscape genetics (Neuenschwander et al., 2008) and in phylogeography where they can be used to assess the current and historical extent of dispersal (Depraz et al., 2008).

For a given model, several simulated genetic datasets can easily be generated by varying parameter values for each simulation. To handle the analysis of such large numbers of simulations efficiently, powerful statistical tools such as approximate Bayesian computations (ABC; Beaumont, Zhang & Balding, 2002; see Table 1) have been developed. ABC statistical procedures offer an efficient and reliable framework for estimating several demographic parameters under complex models and thus inferring demographic histories of populations. Model probabilities, their parameter values as well as other model outcomes can then be compared and assessed using model selection procedures (Johnson & Omland, 2004; Robinson et al., 2013).

The most popular class of genetic data simulators are backward population-based programs that use the coalescent theory (Kingman, 1982; see Table 1) and user-specific models (reviewed in Hoban et al., 2012). These programs allow the generation of different types of genetic data (e.g. SNP, DNA sequences and/or microsatellites) under a wide variety of complex population models and can be easily integrated into the ABC framework via user-friendly software programs for statistical
inference purposes (Cornuet et al., 2008; Lopes, Balding & Beaumont, 2009; Wegmann et al., 2010). In a recent example, Robinson et al. (2013) applied population-based coalescent simulations in an ABC framework to assess the demographic history in a range-restricted and endangered fish species. They tested the effect of an old dam on the genetic diversity of fish populations. Their analysis favoured a constant population size model, demonstrating that the construction of Great Falls Dam on the Caney Fork in Tennessee (U.S.A.) was not responsible for the reduced genetic diversity in the populations sampled, and suggesting that in some cases, conservation efforts should prioritise the maintenance of habitat quality over improving connectivity.

Many population-based coalescent simulators do not consider the spatial complexity of the landscape in which the studied organism lives and thus may not be suitable for answering specific questions about freshwater organisms, particularly in stream systems. To overcome this drawback, Neuenschwander (2006) developed a spatially explicit program (AQUASPLATCHE) that simulates the genetic diversity of populations across semilinear networks such as river systems. By using this program, Neuenschwander et al. (2008) successfully estimated several demographic parameters relative to the colonisation process of the Swiss Rhine basin by the European bullhead Cottus gobio (e.g. migration rates, local effective sizes and the timing of range expansion).

Individual-based forward simulators (Table 1) constitute an important alternative to backward, population-based coalescent simulators, and can be very useful to freshwater ecologists, especially for predictive purposes. To give an example, Chaput-Bardy et al. (2009) used individual-based simulations to examine the effect of river network characteristics (e.g. branching patterns) and dispersal modalities on the genetic differentiation of populations living in dendritic habitats in a virtual landscape. Their results can be used to develop and test hypotheses concerning dispersal in river systems using empirical data. Landguth, Muhlfeld & Luikart (2012) have developed CDFISH, a forward simulator that generates genotypes for individuals living in complex stream systems, allowing the modelling of gene flow on such ecosystems. As more realistic evolutionary models become available and as they are increasingly implemented in user-friendly software, freshwater ecologists will gain access to a better set of tools to test different hypotheses on dispersal and connectivity in stream systems (Csillery et al., 2010; Hoban et al., 2012). This development bears the potential to greatly enhance population genetic (see topics ‘Comparative population structure’ and ‘Population genetics of invasive and managed species’), phylogeographic and also applied conservation genetic research (see topic ‘Conservation genetics’) in freshwater systems.

Population genetics of invasive and managed species

Translocation of species and their introduction to new environments is ongoing, and their impacts present a major topic in biological conservation. Over the last two centuries, many species have been transferred beyond their native ranges. There is an increasing body of molecular ecological research that deals with the genetic aspects of introducing fish of alien species or from non-native populations (Blanchet, 2012). Using molecular approaches, freshwater ecologists can identify the area of origin and invasion routes of an invading species or stock (Thibault, Bernatchez & Dodson, 2009; Estoup & Guillemaud, 2010); estimate effective populations sizes (Table 1) and levels of genetic differentiation of invading and native species (Valiente et al., 2010); assess ecological and evolutionary impacts of invaders and hybridisation with them on the native populations (Nolte et al., 2005; Costedoat et al., 2007; Perrier et al., 2013); detect the presence of an invading species early through environmental DNA-based analysis (Table 1) of water samples (Ficetola et al., 2008; Jerde et al., 2011). Molecular tools are thus allowing invasion biologists to understand more fully the mechanisms behind the invasion and establishment of neobiota, as well as how invasive species affect the genetics and evolution of the native biota.

Assessing the impact of hybridisation between farmed fish stocks and wild fish populations is an important endeavour in this field of research (Perrier et al., 2013). Using a microarray approach with 3557 genes, Roberge et al. (2006) showed that five to seven generations in an altered selection regime (farmed versus wild) were sufficient to generate inheritable changes in gene expression patterns and also a reduction in genetic variability in Atlantic salmon populations. Mixing of farmed and wild populations could thus affect the genetic make-up of wild fish (Hindar et al., 2006) and have detrimental effects (Glover et al., 2004). The effects of stocking farmed Atlantic salmon on wild populations were evaluated by Glover et al. (2012): despite high levels of mixing between farmed and wild populations, the historical population structure persists, but the populations have become genetically more homogenous over 40 years. The authors also find that impacts vary among populations, perhaps in relation to the original population density.
Although the human-mediated mixing of various populations of fish species has been extensive, its impact on intraspecific genetic structure was overlooked for a long time (Kohout et al., 2012). Considering the substantial importance of supplemental stocking and supportive breeding for fisheries and the potential impacts on freshwater ecosystems, this topic should be studied more intensively. For commercially important fishes such as salmonids, carefully worked-out concepts for conserving particular species or local strains may be necessary (Allendorf et al., 2001).

Community ecology using DNA metabarcoding

Molecular methods are increasingly used to identify species in large and complex communities and are rapidly becoming established norms in the study of freshwater biodiversity, habitat assessment and biomonitoring. These methods are replacing the routine morphological identification of 100s or 1000s of specimens, which is time-intensive, potentially error-prone (Haase et al., 2006, 2010) and often provides identification only to genus or family (e.g. for larval benthic invertebrates). An exciting development in freshwater research is the use of these methods in more quantitative ways that can be applied to ecological questions beyond those of community composition.

While DNA barcoding can improve taxonomic resolution and reduces costs and human error in identification, its benefits and caveats remain actively debated among taxonomists and applied stream ecologists (Pfrender et al., 2010; Baird & Sweeney, 2011; DeWalt, 2011; Pilgrim et al., 2011; Sweeney et al., 2011; Baird & Hajibabaei, 2012). Government agencies are actively pursuing the use of DNA barcoding to assess the health of streams and rivers (Hajibabaei et al., 2011; Stein et al., 2013), and important methodological research topics include sampling, preservation, data generation and bioinformatics pipelines (Pfrender et al., 2010). Despite several large-scale efforts, databases of vouchers, identified and sequenced specimens remain far from complete, even in areas where the fauna is well described, and DNA barcoding initiatives have been in place for over a decade. For example, the North American mayfly barcode database was recently tripled to account for ~350 mayfly species, that is, c. 50% of the known fauna in this group (Webb et al., 2012). Importantly, this includes most of the common species, and these could therefore already be identified for biomonitoring purposes using DNA barcodes (Webb et al., 2012). In regions with a rich but largely unknown fauna, for example, many tropical regions, data base development is far behind, but DNA barcoding may promote regional description of biodiversity patterns of the aquatic fauna (Pereira et al., 2013) or ecological studies. For example, Gamboa et al. (2012) used a barcoding approach to identify the gut contents of an aquatic hemipteran (Naucoris sp.) for an assessment of the role of this species in a West African aquatic food web.

While growing databases that include more species and their DNA barcodes would seemingly lead to more accurate matches, there may be an important effect of spatial scale on the accuracy of DNA barcoding. Bergsten et al. (2012) used DNA barcodes to identify water beetles (Dytiscidae) from localities throughout Europe. DNA barcoding was accurate when discerning species within localities and at small geographical scales (10s km); however, accuracy decreased when samples came from larger scales because the distinction between species decreased. The authors argued that the sampling of larger areas was more likely to include closely related species (as opposed to local communities, which contained distant species), thus making individual species appear less distinct. These results suggest that DNA barcoding may provide consistent identifications when a local community is being analysed, but may not be applicable to large-scale studies.

Another breakthrough has been the use of environmental DNA (eDNA; Table 1) methods to assess the DNA from organisms present in the aquatic environment either as dissolved DNA or in cellular debris. This eDNA has been used to detect freshwater mammals, insects, amphibians and fish from freshwater samples (Ficetola et al., 2008; Thomsen et al., 2012). It is necessary to ensure that the procedures developed and applied provide a good balance between cost-savings and taxonomic and ecological accuracy. This will only be achieved by an active collaboration between taxonomists, molecular biologists and applied stream ecologists (Holzenthal et al., 2010; DeWalt, 2011).

DNA barcoding and eDNA increasingly use NGS methodologies to obtain more data and to take advantage of the fact that NGS methods can often deal better with shorter DNA fragments typical of environmental or degraded DNA (Knapp & Hofreiter, 2010), than traditional Sanger-sequencing approaches. While more work is needed before the methods become established standards for detecting the presence and identity of species in a given waterbody or site, an exciting development is their use on addressing more fundamental ecological questions. For example, studies have tested the ability of ‘metabarcoding’ (massively parallel sequencing, where
many sequences can be obtained from a single PCR product, similarly to bacterial cloning) to provide quantitative data based on a relationship between organismal biomass and the number of sequences (Hajibabaei et al., 2012). Such quantitative assessments using DNA barcodes from environmental samples are difficult, because species are not necessarily amplified in proportion to their biomass or abundance; some may be overrepresented and some may be underrepresented or not amplified at all. This ‘PCR bias’ can lead to qualitative and quantitative biases in the results. To avoid using PCR, Zhou et al. (2013) combined a mitochondrial enrichment protocol (Tamura & Aotsuka, 1988) with direct NGS sequencing to analyse a bulk sample of terrestrial arthropods. They were able to recover 97% of the taxa and found a reasonably high correlation between biomass and sequence counts per taxon. These studies suggest that quantitative assessments in environmental barcoding are on the rise.

Conservation genetics

Recent advances in population genetics theory, new developments in technological and analytical tools related to genetics, as well as dramatic reductions of their costs, are leading to new concepts and approaches in freshwater species conservation, freshwater restoration and management (Frankham, 2010; Geist, 2010, 2011). The main contribution of these developments is in increasing the scope and detail of molecular data that can be gathered and applied in studies of freshwater systems.

One outcome of all recent technological advances has been the striking reductions in both the time and the cost required for the development and characterisation of classical markers used in conservation genetics such as microsatellites (Dubut et al., 2010; Malusa et al., 2011; Leese et al., 2012; O’Bryhim et al., 2013). As a result, it is now affordable to perform traditional conservation genetical analyses on many species or on communities/assemblages rather than on a single species (see topic ‘Comparative population structure’). Since many threats to biodiversity, such as habitat fragmentation or climate change, do not affect different species in the same way, multispecies studies are now feasible and becoming more frequent and important for defining management and conservation policies (Davies, Margules & Lawrence, 2000; Blanchet et al., 2010; Gomez-Uchida et al., 2013). In a recent example, Blanchet et al. (2010) used a multispecies comparative approach to reveal species-specific effects of anthropogenic river fragmentation by dams and weirs on the genetic structure of four freshwater fish species and subsequently proposed priority management measures that target the most affected species. Beyond increasing the amount of data, new developments also allow exploitation of different sources of sampling material relevant to species conservation. For example, sequencing of low quality DNA has opened up new possibilities for integrating museum-preserved specimens in conservation research. Bálint et al. (2012) compared mitochondrial sequences of extant populations of a once common mayfly with sequences from museum specimens from its former range, aiming to understand current patterns of genetic diversity of this species, now restricted to ca. 2% of its original area. The study identified the overlaps between historic and current genetic diversity of the species and helped prioritise conservation activities.

Analytical developments have also resulted in valuable new tools that make genetic analyses more applicable to freshwater restoration. For example, methods in landscape genetics have greatly improved the capacity for understanding how landscape elements affect the spatial genetic diversity of populations (Gaggiotti, 2010; Manel & Holderegger, 2013). However, they are mostly used in terrestrial systems and are waiting to be adopted by more freshwater ecologists. For instance, novel Bayesian methods facilitate assessments of the relative effects of multiple geographical and environmental factors (e.g. connectivity, water temperature, altitude) on the genetic structure and/or recent migration of populations (Foll & Gaggiotti, 2006; Faubet & Gaggiotti, 2008). Such tools are very valuable for determining potential threats to freshwater organisms at the river basin scale and may enhance informed decision-making (Leclerc et al., 2008; Cook et al., 2011; Olsen et al., 2011).

Newly developed Bayesian and likelihood-based analyses (see topic ‘Population genetics of invasive and managed species’) have also benefited demographic studies and offer new opportunities for conservation science, allowing for example the detection of declining populations. For instance, the likelihood-based method of Storz & Beaumont (2002), implemented in the software msvar 1.3 (Mark Beaumont, University of Bristol, Bristol, UK), allows effective detection, dating and quantification of demographic changes (e.g. bottlenecks) based on neutral genetic data (Chikhi et al., 2010; Girod et al., 2011; Paz-Vinas et al., 2013a). Paz-Vinas et al. (2013b) used this method to reveal recent human-related bottlenecks on several populations of the endangered freshwater fish Parachondrostoma toxostoma and thus detect the high risk of extinction of these populations.
Such inference methods should, however, be used carefully, since they are generally based on simple demographic models (e.g. Wright–Fisher model) and deviations of real populations from these models may lead to misinterpretations or incorrect inferences (e.g. asymmetric gene flow in rivers; Paz-Vinas et al., 2013a). Important improvements were also made concerning genetic-based estimation of census population size (Nc) and effective population size (Në), allowing freshwater ecologists and wildlife managers to monitor threatened or invasive freshwater populations more efficiently (Luikart et al., 2010; Blanchet, 2012; Palstra & Fraser, 2012). The individual genetic tagging of teleost fish (particularly useful for capture–mark–recapture approaches) has recently been validated and is a further development that will strongly facilitate estimation of fish population sizes in freshwater habitats (Andreou et al., 2012).

Freshwater conservation will also benefit from the increasing integration of genetic diversity data and spatial modelling of species distributions (Bálint et al., 2011; Taubmann et al., 2011; Paz-Vinas et al., 2013b). Assessing whether populations will remain geographically stable or estimating how they may migrate under changing environmental conditions (e.g. using species distribution modelling, SDM; Domisch et al., 2013) should become important drivers for management decisions. Combining these data will permit the prioritisation of populations and regions for conservation that have both a high probability of survival and ideally a higher evolutionary potential (Pfenninger, Bálint & Pauls, 2012; Pauls et al., 2013; Paz-Vinas et al., 2013b).

Much conservation-orientated research in fresh waters relies primarily on neutral genetic diversity. While this is valid and important for assessing population size and demographics, future research on conservation should also identify and monitor loci under selection, that is, loci that may be directly involved in phenotypic variation. Tracking the fate of these loci under changing environmental conditions allows more direct assessments of species vulnerability. With increasing access to large amounts of data – even from a few individuals – this endeavour will become more feasible.

Conclusions

Molecular ecological research in fresh waters has a strong foundation, based on more than 40 years of using genetic tools to study population structure, taxonomy, systematics and evolution. The field has recently undergone a significant transformation, largely owing to advances in technology and in analytical models and methods. The possibilities of applying these new approaches to freshwater biology are many, but here, we identified eight topics that we find particularly promising. These span a continuum from basic to applied research and range from local adaptation to large-scale evolution. A common theme is the benefits to be gained from an increasing ability to carry out multispecies, multilocus studies to test for general patterns. Distinct advantages include the study of non-model organisms, natural populations, and functionally and ecologically relevant genetic markers. New advances also provide the means to direct the study of environment–phenotype–genotype interactions towards genomes and transcriptomes. Linking them with experimental work will improve our understanding of adaptations in freshwater species, and how they have variously evolved from terrestrial and marine ancestors. The increasing amount of available data requires new approaches of data handling and analysis. Care must be taken to not discard recent conceptual and analytical developments, such as approaches based on simulation or the merger of population genetic and community ecology, in a tradeoff for simply obtaining more data. Finally, the integration of new developments should aspire to strengthen freshwater management and conservation in addition to undertaking basic research.

New technologies can have positive and negative influences on any particular field of research. A danger of the ever-developing technology is that it results in a highly specialised area of inquiry with its own research questions and necessary skill sets. One of our aims here was to consider a more positive alternative, namely that methodological changes will facilitate the increased collaboration of freshwater researchers with a variety of perspectives. This is because of the increasing breadth of taxa and genetic markers to which molecular ecology methods can be applied. We anticipate greater synergy with other approaches, such as behavioural studies, stable isotope analyses, or laboratory and field experiments. The result will be a far better understanding of ecological and evolutionary processes in fresh waters than would be possible by any discipline alone.

Acknowledgments

The manuscript largely reflects the views arising from discussions held among the participants of a Special Session at the 7th Symposium for European Freshwater Sciences (SEFS) held in Girona, Spain in 2011 and again at the 8th SEFS held in 2013 in Münster, Germany. We thank all participants of those two sessions for the
fruitful discussions. We thank two anonymous reviewers and Alan Hildrew for constructive comments that helped improve the manuscript. SUP’s research is supported by the research funding programme ‘LOEWE – Landes-Offensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz’ of Hesse’s Ministry of Higher Education, Research, and the Arts. I. P-V. works in two research units (EDB & SEEM) that are part of the ‘Laboratoire d’Excellence’ (LABEX) entitled TULIP (ANR-10-LABX-41).

References


start-feeding of brown trout families. *Journal of Fish Biology, 64*, 1168–1178.


© 2014 John Wiley & Sons Ltd, Freshwater Biology, 59, 1559–1576.


(Manuscript accepted 30 March 2014)