UC Merced
Frontiers of Biogeography

Title
Massively parallel DNA sequencing: the new frontier in biogeography

Permalink
https://escholarship.org/uc/item/7xv39142

Journal
Frontiers of Biogeography, 5(1)

Authors
Rocha, Luiz A.
Bernal, Moisés A.
Gaither, Michelle R.
et al.

Publication Date
2013

DOI
10.21425/F5FBG16189

Copyright Information
Copyright 2013 by the author(s).This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

eScholarship.org Powered by the California Digital Library
University of California

perspective

ISSN 1948-6596

Massively parallel DNA sequencing: the new frontier in biogeography

Luiz A. Rocha[1,*], Moisés A.Bernal[1,2], Michelle R. Gaither[1] and Michael E. Alfaro[3]

[1]California Academy of Sciences, Section of Ichthyology, 55 Music Concourse Drive, San Francisco, CA 94118, USA

[2]University of Texas Marine Science Institute, 750 Channel View Drive, Port Aransas, TX 78373, USA

[3]Department of Ecology and Evolutionary Biology, 2154 Terasaki Life Sciences Building, University of California, Los Angeles, CA 90095, USA

*LRocha@calacademy.org; http://research.calacademy.org/ichthyology/

Abstract. The advent of Sanger sequencing represented a scientific breakthrough that greatly advanced biogeographic studies. However, this technology has several limitations that have hampered more advanced studies in the field. The development of novel techniques which more fully exploit the potential of Massively Parallel Sequencing (MPS) to deliver sequence data at a fraction of the cost of Sanger sequencing promises to revolutionize biogeographic studies. Approaches like Restriction-site Associated DNA sequencing (RADseq) and UltraConserved Element (UCE) sequencing enable the collection of unprecedented amounts of data for multi-locus studies of population genetics and phylogenetics respectively, which in turn can be used for biogeographic analysis. Here we review those and other methods related to MPS, and provide examples of how they can be used in tropical Atlantic biogeography.

Keywords. next-generation sequencing, phylogeography, phylogenetics, genomics, hybridization

Introduction

Traditionally, biogeography has relied upon species distributions to understand what factors shape present and historical biodiversity (Ekman 1953, Briggs 1974). The advent of cladistics provided a new tool for the biogeographer, as the comparison between geographical and biological cladograms, and their inferred histories, allowed for the first time the analysis of congruence between those two sets of data and the potential discovery of geographic processes that shape species distributions (Rosen 1978). The advent of PCR-based DNA sequencing accelerated biogeographic analyses by providing new tools and methods designed to test biogeographic hypotheses in the context of molecular phylogenies and geographic distribution of genetic lineages (Avise 2000, Lomolino and Heaney 2004, Floeter et al. 2008, Briggs and Bowen 2012).

Although first-generation sequencing technologies (Sanger et al. 1977) greatly advanced biogeographic studies, relative to MPS, DNA marker development and sequencing under this technology is labor intensive and expensive. As a result, most biogeographic studies are based upon the analysis of a small number of loci (Shendure and Ji 2008). While the sequencing of a handful of genes has significantly contributed to our understanding of evolutionary relationships and biogeographical processes, filling in the data gaps and resolving intractable phylogenies and phylogeographies will require data sets with more loci, and greater taxonomic coverage. The analysis of multiple unlinked DNA loci has the potential to increase our confidence in reconstructing species histories, resolving taxon boundaries, analyzing modes of speciation, and determining evolutionary relationships (Wu and Ting 2004, Dupuis et al. 2012, Hohenlohe et al. 2012). Massively parallel DNA sequencing (MPS) platforms produce enormous amounts of sequence data and the development of methods for tagging and pooling genomic libraries has made possible the multiplexing of hundreds of samples (Emerson et al. 2010, Faircloth et al.

This manuscript is part of the proceedings of the *Workshop on the Biogeography and Phylogeography of Atlantic Fish* (Lisbon, November 2011).

© 2013 the authors; journal compilation © 2013 The International Biogeography Society

2012). These new sequencing technologies are dramatically increasing the efficiency of genomic data acquisition and lowering costs are rendering MPS feasible for most labs (McCormack et al. 2013a), however, bioinformatics (both in terms of processing power and programming) remains as an important limitation.

Here we describe how MPS promises to change the field of biogeography. We will not review sequencing platforms, methods, or analyses as these have been reviewed extensively elsewhere (Shendure and Ji 2008, Kircher and Kelso 2010, Metzker 2010, McCormack et al. 2013a). Instead, we focus on the applications of MPS to the study of underlying causes of organismal distributions. We highlight cases where MPS platforms have been employed to improve the resolutions of biogeographic questions and consider how the advent of MPS will impact biogeography studies in general and Atlantic studies in particular.

Inferring phylogenies

Resolving the evolutionary relationships between species and the historic events that drive patterns of speciation is one of the cornerstones of biogeographic studies (Ronquist and Sanmartin 2011). Molecular systematics has traditionally relied upon sets of a few genes representing a tiny fraction of the genome to resolve phylogenies. Locus discovery has been a limiting step for many projects due to the time and effort required for primer design, tests to ensure cross species amplification, and optimization of PCRs. Second, and now third-generation sequencing technologies are rapidly changing this paradigm, allowing the collection of millions of base pairs of data or even entire genomes, for non-model organisms (Glenn 2011). The push for larger numbers of independent loci is driven in part by the recognition that individual gene trees are often discordant (Maddison 1997), a fact that has resulted in a move toward analytical approaches that attempt to reconcile discordant gene trees within a single phylogeny (Degnan and Rosenberg 2009, Edwards 2009). Further driving the field are developments in analytical software that make analyses of large multi-locus data sets possible (Hey and Nielsen 2007, Heled and Drummond 2010, Yang and Rannala 2010).

Amplicon sequencing

Several methods use MPS platforms to sequence sets of orthologous loci across taxa. Choosing a suitable technique depends on number of loci desired, bioinformatic capability available, and whether a specific set of loci of known function is desired. Some phylogenetic studies would benefit from the sequencing of specific genes or the use of a set of known loci across a large number of individuals. This targeted sequencing of previously developed orthologous loci can be accomplished by using a class of methods termed amplicon sequencing or parallel tagged sequencing (Meyer et al. 2008). This method involves the pooling of tagged PCR products from multiple individuals and is particularly useful when a collection of known markers, which reliably amplifies across the taxonomic group of interest, is available (Bybee et al. 2011). The advantage of amplicon sequencing over traditional Sanger sequencing is that data for large numbers of individuals at a few to dozens of loci can be sequenced quickly at low cost. Also, because MPS data are single stranded, issues involving allelic phase for nuclear loci are resolved, doing away with the need for expensive and time-consuming cloning. However, these methods do not eliminate several of the labor and cost intensive aspects of Sanger sequencing: PCRs are needed for each individual at each locus, and locus discovery and primer design remain as potentially time consuming steps (Table 1). Multiplex PCRs could circumvent some of the front-end laboratory work, however, these methods have their problems such as the production of spurious amplification products and uneven amplification of some target sequences (Elnifro et al. 2000). Avoiding these pitfalls can require time-consuming optimization experiments.

Despite the limitations, amplicon sequencing has been a successful strategy in many mitochondrial genome based studies. Chan et al. (2010) sequenced 51 mitogenomes across three genera of *Hylobates* gibbons, producing a robust

 © 2013 the authors; journal compilation © 2013 The International Biogeography Society

phylogeny. Morin et al. (2010) were able to resolve previously intractable polytomies among ecotypes of Killer Whales (*Orcinus orca*) using the full 16,390 base pair mitogenomes of 143 individuals. Amplicon sequencing has shown some utility in sequencing large numbers of nuclear loci, though published studies to date have been primarily of methodological nature. For instance, in a test case, Bybee et al. (2011) sequenced six markers from 44 taxa of crustaceans in one quarter of a plate on a Roche 454 and O'Neill et al. (2013) used similar methods to sequence 95 nuclear loci in 93 individuals of the tiger salamander (*Ambystoma tigrinum*) species complex, but both studies are heavily focused on describing their methods.

Targeted enrichment

Targeted enrichment of DNA (Hodges et al. 2007) ‘‘captures’’ specific DNA sequences by hybridizing target DNA to probes in solution or anchored to a surface (i.e., beads, microarray). Recent approaches that capitalize on this technology make use of a class of markers that are anchored by ultra-conserved regions of the genome. The function of these ultra-conserved elements (UCEs) is not yet understood but they are seemingly ubiquitous across animals and due to their highly conserved nature, they are ideal regions for probe design. A method described by Faircloth et al. (2012) allows sequencing of thousands of loci without locus-specific PCR. Probes are designed by screening alignments of published genomes from divergent taxa for regions of high sequence similarly. During library preparation, the genomic DNA is fragmented, mixed with and allowed to hybridize to the DNA (or RNA) probes. The “captured” DNA is amplified and then sequenced using Illumina technology. Variation in the regions flanking the UCEs provide thousands of independent orthologous loci distributed across the genome and because of high conservation these loci are easy to align across divergent taxa making UCEs useful for phylogenetic studies (Stephen et al. 2008). In a series of studies, UCEs were employed to recover established phylogenies among basal lineages of birds (McCormack et al. 2013b), to resolve the placement of turtles as sister to birds and crocodilians (Crawford et al. 2012), and to resolve the relationships among placental mammals (McCormack et al. 2012), showing that

Methods	Application	Benefits	Limitations
Amplicon sequencing	Phylogenetics and phylogeography studies that require a limited number of loci	Because the loci are targeted with specific primers it allows easy reproducibility across samples and labs	Requires individual PCRs for each sample at each locus. Labor intensive and cost prohibitive for large-scale studies
Target Enrichment (UCEs, Anchored Enrichment, EPIC)	Phylogenetics and phylogeography that require a large number of loci	Recovers thousands of loci simultaneously. Conserved regions can recover phylogenies among basal lineages	Locus design is time consuming and requires extensive genomic resources. Results may be biased if taxa are too distant from the reference probes
Restriction Enzymes (Rad-tags and Genotype by Sequencing)	Phylogeography, population genetics, detecting selection	Very even sampling across the genome that can be done without previous genome knowledge of the group of interest. Good for very recent timescales	Limited reproducibility across different runs. Restriction sites may vary considerably between species, which may limit the number of loci available for comparisons
Transcriptome (RNA-seq)	Gene expression, selection, phylogenetics and recently applied to population genetics	Because of its conserved nature, allows for the identification of gene function even when genomic resources are scarce	Tissues require special care before extractions. Highly expressed genes may dominate the sequences

Table 1. Commonly used genomic methods for phylogenetics and phylogeography of non-model organisms, their applications, benefits and limitations.

 © 2013 the authors; journal compilation © 2013 The International Biogeography Society

this technique holds great promise for resolving deep level phylogenies. Lemmon et al. (2012) describe a similar method for phylogenetic reconstruction termed anchored enrichment, however, these authors amplify fewer loci (tens to hundreds) and target less conserved regions for probe design, both measures that allow for the sequencing of more individuals to resolve questions at theoretically shallower time-scales than UCEs. The reported advantage of targeted enrichment methods is the ability to adjust the time scale by increasing target DNA length (longer fragments should equal more variation) and once a capture probe set is developed for a taxonomic group the same set of probes can be used to obtain loci from any collection of organisms within that group. However, similar to amplicon sequencing, locus design can be time consuming (Table 1), but probe sets are already available for many taxa, including reptiles, birds and mammals.

Transcriptomes

RNA molecules (the transcriptome) are another source of orthologous loci and offer several advantages. Transcriptomes are by nature a reduced representation of the genome, making sequencing less costly and faster than whole genome sequencing (Table 1). The availability of published transcriptomes and the presence of fewer repetitive elements in transcripts make their assembly easier than whole genomes (Grabherr et al. 2011).

Transcriptome studies begin with RNA extraction, reverse transcription of RNA into cDNA, and DNA sequencing for the taxonomic group of interest. Bi et al. (2012) described a modified sequence capture technique in which the sequenced transcriptome is assembled and annotated (using published transcriptomes) and specific exons are chosen for array design. However, because these arrays are designed from taxon specific transcriptomes they may be most useful in sets of taxa of low to moderate phylogenetic distances and perhaps even for population genetics and demography studies. In contrast, Hittinger et al. (2010) forgone array design and instead directly sequenced cDNA libraries from species of *Anopheles* mosquito. The authors demonstrated that hundreds of orthologous genes could be consistently recovered using this strategy, and that these gene sets enable robust phylogenomic analysis. Similar approaches have been applied to deeper level phylogenetic questions involving mollusks (Kocot et al. 2011, Smith et al. 2011b) and the placement of the enigmatic Myzostomida marine invertebrates in the tree of life (Hartmann et al. 2012).

An advantage brought by transcriptome analysis, especially for low-level phylogenetics, is that the RNA being analyzed is all functional, and therefore may be affected by selection. This allows researchers to not only detect outlier loci that might be under selection, but may also reveal the function of those loci, which in turn can be tied to environmental selective pressures (Schwarz et al. 2009). However, transcriptome analyses have one important limitation: tissue sample quality. RNA degrades much faster than DNA, and tissue samples collected years ago that could be used for genomic DNA analysis are almost completely devoid of RNA. Therefore, with few exceptions, fresh tissue collections and special preservation methods are needed for transcriptome studies (Copois et al. 2007).

Whole genomes

As the cost of MPS continues to fall, whole genome sequencing is becoming more feasible for a larger number of laboratories. Complete genome sequencing is now possible for non-model organisms providing an opportunity for genome-wide comparisons to examine structural genomic variation and recombination, chromosomal rearrangements and variations in gene copy number. However, inadequate bioinformatic capacity (both hardware and software) remains the main impediment for large scale whole genome comparisons, and the few studies using this methodology are still limited to model organisms and use only a subset of the data (e.g., *Drosophila*; Begun et al. 2007).

Defining species boundaries and investigating hybrid zones

Defining species boundaries has long been a con-

 © 2013 the authors; journal compilation © 2013 The International Biogeography Society

tentious topic in evolutionary biology (Sites and Marshall 2003). Over the past several decades, the use of DNA sequences, rather than exclusive reliance on morphology, has become increasingly common in taxonomic studies. However, this use remains somewhat controversial, even when cryptic genetic variation is discovered (Knowlton 2000, Paquin and Hedin 2004). This is especially true for the now widespread practice of using one or a few regions of the genome (DNA barcoding) to diagnose species (Moritz and Cicero 2004). Critics of this barcoding point out that 1) genetic markers commonly used to resolve species boundaries have variable power of resolution across different taxonomic groups (Hollingsworth 2011), 2) the commonplace reliance on a single locus ignores introgression and can lead to erroneous identifications (Croucher et al. 2004, Rocha et al. 2008b, DiBattista et al. 2012) and 3) the low mutation rates of most DNA barcoding genes will likely underestimate biodiversity (Elias et al. 2007). Traditional barcoding methods are likely to be suitable for identifying unknown individuals of well-resolved taxa (e.g., unidentified fish fillets or larvae; Wong and Hanner 2008, Weigt et al. 2012) and, despite their limitations, have been instrumental in the detection of distinct cryptic lineages among morphologically conserved species (Bickford et al. 2007, Burns et al. 2008). However, the small size of most first-generation sequencing datasets often precludes robust resolution of species boundaries in cases where gene flow is ongoing or differentiation is very recent (e.g., African cichlids; Wagner et al. 2013). Even though inconsistencies will probably remain in some cases, the development of MPS-based approaches promises to overcome many of these limitations through the sheer number of loci that can be economically interrogated.

Hybridization

The study of zones in which species hybridize helps us understand how speciation can proceed with gene flow (Hewitt 2008, Payseur 2010). The post-glacial suture zones of North America and Europe are model examples, where climate warming and glacial melting following the end of the last glacial maximum allowed populations to expand, bringing previously isolated species back into secondary contact and leading to hybridization (Hewitt 1999, Swenson and Howard 2005). For example, extensive introgression and replacement of polar bear mtDNA with brown bear mtDNA was detected (Edwards et al. 2011), and this evidence suggested a time to most recent common ancestor (TMRCA) of 150kya between hybridizing populations (Lindqvist et al. 2010). A recent study applying MPS data to this system (Miller et al. 2012) illustrates the power of this approach to enrich phylogeographic studies. More than one million single nucleotide polymorphisms (SNPs) of nuclear DNA were used to perform phylogenetic analyses between polar, black and brown bears (Miller et al. 2012). These results confirm the hybridization, but also indicate that polar bears and brown bears initially split 4-5 Mya, experienced a long interval with little to no gene flow, and hybridized only recently, likely because of the effects of historical climatic fluctuations. Furthermore, the large number of markers allowed the construction of a detailed time line of the effective population size of polar bear populations through the application of a pairwise sequential Markovian coalescent model (Li and Durbin 2011). This model estimates TMRCA in diploid sequences using changes of density of heterozygous sites across the genome, and indicates that the effective population size (*Ne*) of this species is strongly associated with climatic oscillations (Miller et al. 2012).

It has long been assumed that hybrid zones were not as common in marine systems due to the high dispersal potential of marine organisms. However, the recent discovery of hybrid reef fishes at Christmas and Cocos Keeling islands (eastern Indian Ocean) highlighted the commonness and importance of hybridization across several marine fish families (Hobbs et al. 2009). It is hypothesized that changes in sea level during climatic oscillations led to the exposure of a land bridge between the Pacific and Indian Ocean around the Coral Triangle area, causing differentiation through isolation between species in these two oceans (Woodland 1983, Gaither and Rocha

© 2013 the authors; journal compilation © 2013 The International Biogeography Society

2013). Once sea levels increased, taxa that were separated came into secondary contact, especially around the islands of Christmas and Cocos Keeling, just west of Indonesia. Today, 11 pairs of reef fishes of six different families of the Pacific and Indian Ocean are known to hybridize in coral reefs around those islands (Hobbs et al. 2009). Since hybridization is proving to be much more common in marine fishes than previously thought, the unprecedented ability of recovering thousands of nuclear loci for non-model organisms offers the opportunity to study speciation with gene flow and reticulate evolution on a large scale (Bowen et al. 2013).

Another example where MPS provided new insights on hybridization is in the diverse *Heliconius* butterflies. Several species of *Heliconius* are unpalatable and Müllerian mimicry of warning colors enables species to share the cost of educating predators. Multilocus gene genealogies among three species of this genus (two with overlapping distributions, and the third with a distinct range) revealed a striking pattern: gene trees related to mimetic color patterns show a closer relationship between species with sympatric distribution, whereas other genes put two of the allopatric species together. Thus, introgression in this group plays an adaptive role, as the genes with the strongest signal of introgression are those that confer the advantage of Müllerian mimicry (Dasmahapatra et al. 2012, Pardo-Diaz et al. 2012).

Morphological conservatism

The advent of DNA sequencing raised awareness about cryptic species, as it became easy for researchers to detect and differentiate unique genetic lineages within morphologically very similar or even identical species (Bickford et al. 2007). However, species delimitations are usually done with one or a few loci, which does not allow the exploration of the genomic mechanisms underlying morphological conservatism. The mechanisms driving such conservatism are still unknown. Traditionally, three non-mutually exclusive hypotheses are used to explain cases in which single DNA markers show divergence but morphology does not: random mutation and genetic drift that does not affect morphological characters; developmental constraints limiting the evolution of phenotypes; and similar selective pressures in similar environments that result in conserved morphology (Arnold 1992, Smith et al. 2011a). Considering the lowering costs and successful applications of MPS in non-model organisms, it is now feasible to track the evolutionary history of thousands of individual loci (Hohenlohe et al. 2012), develop linkage maps in only one generation (Amores et al. 2011), and determine the role of loci differentially expressed between lineages within individual groups (Baldo et al. 2011). Even though many cases where cryptic species are discovered may not represent true morphological stasis and differences might be revealed by more detailed morphological analyses (Knowlton 2000), MPS has the potential to bring us a much better understanding of cryptic speciation.

Phylogeographic surveys

The vast majority of phylogeographic studies are based upon analysis of one or several tightly linked mitochondrial markers (Avise 2000, Rocha et al. 2007). This bias is a response to the limited availability of primers that can be used across large numbers of taxa, and their presumed neutrality and fast mutation rates (Avise 2000). The limitations of this approach are now well appreciated. For example, hybridization commonly leads to introgression, masking any signal of divergence between closely related species (Rocha et al. 2008b). Further, because mtDNA is maternally inherited, results may be skewed when females and males have different life histories, (e.g., male mediated gene-flow in sea turtles; Bowen and Karl 2007). The more recent trend of including nuclear DNA markers represents an important development in phylogeography.

Single Nucleotide Polymorphisms

Microsatellites have been favored as the standard for population genetics with traditional sequencing methods, however, single nucleotide polymorphisms (SNPs) offer a powerful alternative (Coates et al. 2009). On the downside, SNPs have a lower

 © 2013 the authors; journal compilation © 2013 The International Biogeography Society

power of resolution: the power given by 10 to 20 microsatellites has been estimated to be equivalent to the use of 100 SNPs (Liu et al. 2005). In addition, considering that *F* statistics are calculated over all loci, divergence can be overestimated when some of the observed loci are under selection pressure, or linked to such regions (Helyar et al. 2011). With the advent of MPS both of these issues can now be easily resolved, as tens of thousands of SNPs can be recovered via the use of Restriction Site Associated DNA (RAD) methods for multiple individuals at the same time. This methodology reduces the representation of the genome by cutting it in predetermined sizes and orthologous positions, but the large amount of restriction site position variation above the species level limits this technique to population or closely related species level studies (Baird et al. 2008, Peterson et al. 2012, Table 1). Analyzing patterns of population structure from thousands of loci obtained in a single run can overwhelm signals of selection caused by a small number of outlier loci (Emerson et al. 2010) or help us detect outlier loci under the influence of selection (Hohenlohe et al. 2012). This may lead to the replacement of microsatellites by SNPs, as the main marker for the study of population genetics in the near future (Seeb et al. 2011).

Even though loci under strong divergent selection might inflate the level of differentiation between populations, they are becoming a useful tool for individual assignment tests between fisheries stocks that have low levels of differentiation with traditional markers. For example, the populations of salmon from Eastern Canada have been subdivided in stocks of the Inner and outer Bay of Fundy, based on their migratory behavior. Despite the presence of rare mitochondrial haplotypes unique to the Inner Bay, there is limited evidence of genetic differentiation between the two locations, as previous studies have not been able to differentiate members of both populations (Fraser et al. 2007). However, using 320 SNPs, Freamo et al. (2011) were able to find small but significant differences between those populations. Further, four independent tests demonstrated that eight of the observed loci were under divergent selection. In a separate test, nine loci were genotyped through an assay in order to perform assignment tests. The assignment test was concordant with the SNP matrix, as it revealed small but significant divergence between individuals of the inner and outer Bay of Fundy (Freamo et al. 2011). Considering that in recent years there has been a significant decline in the number of adults that return to spawn at the Inner Bay, this tool is very useful for conservation as it enables the determination of the origin of the few adults that reach spawning sites.

MPS and tropical Atlantic biogeography

The tropical Atlantic represents an excellent system in which to test biogeographic and evolutionary hypotheses. It is relatively small in size when compared to the Indo-Pacific and somewhat closed, bordered to the east and west by continental landmasses and to the north and south by cold waters. This relative simplicity has allowed for detailed studies and syntheses about biogeographic processes at work in the Atlantic (reviewed in Floeter et al. 2008). But all of this work has been based either on phylogenies or phylogeographies estimated with Sanger sequencing technology or species distributions, so a crucial question is: which advances can MPS bring to our understanding of diversification patterns in the Atlantic?

One possibility is the use of MPS to compare the signal left in the genome by different speciation modes. In addition to the characteristics mentioned above, the Atlantic was also the stage of a dramatic geologic event, the closure of the Isthmus of Panama. When the Isthmus closed, several species were divided in two, and since the geology of the area is well known, we can date back the time of separation between these pairs to at least 3 My (Lessios 2008). There are also many examples of possible selection-driven speciation with gene flow in Atlantic fishes (Rocha et al. 2005a, Puebla et al. 2007), and some genera contain possible examples of both cases (Rocha et al. 2008b), which provide unique opportunities for comparisons using MPS. The use of SNP analyses through RAD-sequencing can quickly provide

 © 2013 the authors; journal compilation © 2013 The International Biogeography Society

much needed additional information to confirm these proposed cases of speciation with gene flow, and when coupled with comparisons of species pairs that diverged in isolation, they might reveal a genomic signature of speciation mode. MPS can also bring better resolution to analyses of events that occurred in a more recent time scale. Several lines of evidence indicate that two of the most important biogeographical barriers in the Atlantic, the Amazon barrier and the Benguela barrier, are porous and their effectiveness changes with sea level and climatic fluctuations: the Amazon is frequently breached in periods of low sea level, and Benguela is crossed during warm interglacials (Rocha 2003, 2005b, 2008a, Robertson et al. 2006). Through RAD-sequencing, thousands of loci can be obtained to estimate migration rates and conduct coalescent analysis with higher confidence than the current estimates based on just a few genes. However, this method suffers from similar limitations as transcriptomes and works better with recently preserved tissues.

On the phylogenetic side, relationships among species of many important groups remain either completely or partially unresolved. Many Atlantic species are missing from phylogenetic hypotheses of several important coral reef fish groups (Rüber et al. 2003, Barber and Bellwood 2005, Fessler and Westneat 2007, Miller and Cribb 2007), and trees are not even presented for other important groups such as the basses (fishes of the family Serranidae). The few complete (or close to complete) phylogenies presented (Craig and Hastings 2007, Sanciangco et al. 2011) could be better resolved with more loci. Currently, the best tool to resolve phylogenies is the sequencing of areas adjacent to UCEs, which are being used with success in several other groups (McCormack et al. 2013b). But regardless of which tool is used to answer which question, one thing is certain, MPS will revolutionize the way we do biogeography, and we think the changes are welcome as we will finally be able to explore many questions previously impossible to properly address with current tools.

Acknowledgements

We thank Brian Simison and Durrell Kapan for lively discussions and intellectual input. We also thank Vitor Almada, Rita Castilho, and André Levy for organizing the workshop on the biogeography and phylogeography of Atlantic fishes in Lisbon, where the idea to write this manuscript was sparked. The workshop took place in November of 2011, in Lisbon, Portugal, and during that time we had lively discussions about the role that MPS will play in future developments in biogeography. Financial support to attend the workshop was provided by the California Academy of Sciences.

References

Amores, A., Catchen, J., Ferrara, A., Fontenot, Q. & Postlethwait, J.H. (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: Spotted Gar, an outgroup for the teleost genome duplication. Genetics, 188, 799–808.

Arnold, S.J. (1992) Constraints on phenotypic evolution. The American Naturalist, 140 S1, S85–S107.

Avise, J.C. (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, 3, e3376.

Baldo, L., Santos, M.E. & Salzburger, W. (2011) Comparative transcriptomics of Eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. Genome Biology and Evolution, 3, 443–455.

Barber, P. & Bellwood, D.R. (2005) Biodiversity hotspots: evolutionary origins of biodiversity in wrasses (*Halichoeres*: Labridae) in the Indo-Pacific and new world tropics. Molecular Phylogenetics and Evolution, 35, 235–253.

Begun, D.J., Holloway, A.K., Stevens, K., et al. (2007) Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. PLoS Biology, 5, e310.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C. & Good, J.M. (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics, 13, 403.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K. & Das, I. (2007) Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution, 22, 148–155.

Bowen, B.W. & Karl, S.A. (2007) Population genetics and phylogeography of sea turtles. Molecular Ecology, 16, 4886–4907.

Bowen, B.W., Rocha, L.A., Toonen, R.J., et al. (2013) The origins of tropical marine biodiversity. Trends in Ecology and Evolution, in press.

 © 2013 the authors; journal compilation © 2013 The International Biogeography Society

Briggs, J.C. (1974) Marine Zoogeography. McGraw-Hill, New York.

Briggs, J.C. & Bowen, B.W. (2012) A realignment of marine biogeographic provinces with particular reference to fish distributions. Journal of Biogeography, 39, 12–30.

Burns, J.M., Janzen, D.H., Hajibabaei, M., Hallwachs, W. & Hebert, P.D.N. (2008) DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. Proceedings of the National Academy of Sciences USA, 105, 6350–6355.

Bybee, S.M., Bracken-Grissom, H., Haynes, B.D., Hermansen, R.A., Byers, R.L., Clement, M.J., Udall, J.A., Wilcox, E.R. & Crandall, K.A. (2011) Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biology and Evolution, 3, 1312–1323.

Chan, Y.C., Roos, C., Inoue-Murayama, M., Inoue, E., Shih, C.C., Pei, K.J. & Vigilant, L. (2010) Mitochondrial genome sequences effectively reveal the phylogeny of Hylobates gibbons. PLoS ONE, 5, e14419.

Coates, B.S., Sumerford, D.V., Miller, N.J., Kim, K.S., Sappington, T.W., Siegfried, B.D. & Lewis, L.C. (2009) Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. Journal of Heredity, 100, 556–564.

Copois, V., Bibeau, F., Bascoul-Mollevi, C., et al. (2007) Impact of RNA degradation on gene expression profiles: Assessment of different methods to reliably determine RNA quality. Journal of Biotechnology, 127, 549–559.

Craig, M.T. & Hastings, P.A. (2007) A molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelini. Ichthyological Research, 54, 1–17.

Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K. & Glenn, T.C. (2012) More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters, 8, 783–786.

Croucher, P.J.P., Oxford, G.S. & Searle, J.B. (2004) Mitochondrial differentiation, introgression and phylogeny of species in the *Tegenaria atrica* group (Araneae : Agelenidae). Biological Journal of the Linnean Society, 81, 78–89.

Dasmahapatra, K.K., Walters, J.R., Briscoe, A.D., et al. (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature, 487, 94–98.

Degnan, J.H. & Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology and Evolution, 24, 332–340.

DiBattista, J.D., Waldrop, E., Bowen, B.W., Schultz, J.K., Gaither, M.R., Pyle, R.L. & Rocha, L.A. (2012) Twisted sister species of pygmy angelfishes: discordance between taxonomy, coloration, and phylogenetics. Coral Reefs, 31, 839–851.

Dupuis, J.R., Roe, A.D. & Sperling, F.A.H. (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. Molecular Ecology, 21, 4422–4436.

Edwards, C.J., Suchard, M.A., Lemey, P., et al. (2011) Ancient hybridization and an Irish origin for the modern polar bear matriline. Current Biology, 21, 1251–1258.

Edwards, S.V. (2009) Is a new and general theory of molecular systematics emerging? Evolution, 63, 1–19.

Ekman, S. (1953) Zoogeography of the Sea. Sidgwick and Jackson, London.

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Malllet, J. & Jiggins, C.D. (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proceedings of the Royal Society B, 274, 2881–2889.

Elnifro, E.M., Ashshi, A.M., Cooper, R.J. & Klapper, P.E. (2000) Multiplex PCR: optimization and application in diagnostic virology. Clinical Microbiology Review, 13, 559–570.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E. & Holzapfel, C.M. (2010) Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences USA, 107, 16196–16200.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. & Glenn, T.C. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology, 61, 717–726.

Fessler, J.L. & Westneat, M.W. (2007) Molecular phylogenetics of the butterflyfishes (Chaetodontidae): Taxonomy and biogeography of a global coral reef fish family. Molecular Phylogenetics and Evolution, 45, 50–68.

Floeter, S.R., Rocha, L.A., Robertson, D.R., et al. (2008) Atlantic reef fish biogeography and evolution. Journal of Biogeography, 35, 22–47.

Fraser, D., Jones, M., McParland, T. & Hutchings, J. (2007) Loss of historical immigration and the unsuccessful rehabilitation of extirpated salmon populations. Conservation Genetics, 8, 527–546.

Freamo, H., O'Reilly, P., Berg, P.R., Lien, S. & Boulding, E.G. (2011) Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Molecular Ecology Resources, 11, 254–267.

Gaither, M.R. & Rocha, L.A. (2013) Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the Centre of Overlap hypothesis. Journal of Biogeography, accepted

Glenn, T.C. (2011) Field guide to next-generation DNA sequencers. Molecular Ecology Resources, 11, 759–769.

Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29, 644–652.

Hartmann, S., Helm, C., Nickel, B., Meyer, M., Struck, T.H., Tiedemann, R., Selbig, J. & Bleidorn, C. (2012) Exploiting gene families for phylogenomic analysis of myzostomid transcriptome data. PLoS ONE, 7, e29843.

Heled, J. & Drummond, A.J. (2010) Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution, 27, 570–580.

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., et al. (2011) Application of SNPs for population genetics of non-model organisms: new opportunities and challenges. Molecular Ecology Resources, 11, 123–136.

© 2013 the authors; journal compilation © 2013 The International Biogeography Society

Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society, 68, 87–112.

Hewitt, G.M. (2008) Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Molecular Ecology, 10, 537–549.

Hey, J. & Nielsen, R. (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proceedings of the National Academy of Sciences USA, 104, 2785–2790.

Hittinger, C.T., Johnston, M., Tossberg, J.T. & Rokas, A. (2010) Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. Proceedings of the National Academy of Sciences USA, 107, 1476–1481.

Hobbs, J.-P.A., Frisch, A.J., Allen, G.R. & Van Herwerden, L. (2009) Marine hybrid hotspot at Indo-Pacific biogeographic border. Biology Letters, 5, 258–261.

Hodges, E., Xuan, Z., Balija, V., et al. (2007) Genome-wide in situ exon capture for selective resequencing. Nature Genetics, 39, 1522–1527.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. & Cresko, W.A. (2012) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics, 6, e1000862.

Hollingsworth, P.M. (2011) Refining the DNA barcode for land plants. Proceedings of the National Academy of Sciences USA, 108, 19451–19452.

Kircher, M. & Kelso, J. (2010) High-throughput DNA sequencing – concepts and limitations. BioEssays, 32, 524–536.

Knowlton, N. (2000) Molecular genetic analyses of species boundaries in the sea. Hydrobiologia, 420, 73–90.

Kocot, K.M., Cannon, J.T., Todt, C., et al. (2011) Phylogenomics reveals deep molluscan relationships. Nature, 477, 452–456.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. Systematic Biology, 61, 727–744.

Lessios, H.A. (2008) The great American schism: divergences of marine organisms after the rise of the Central American isthmus. Annual Review of Ecology, Evolution, and Systematics, 39, 63–91.

Li, H. & Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. Nature, 475, 493–496.

Lindqvist, C., Schuster, S.C., Sun, Y., et al. (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. Proceedings of the National Academy of Sciences USA, 107, 5053–5057.

Liu, N.J., Chen, L., Wang, S., Oh, C.G. & Zhao, H.Y. (2005) Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. BMC Genetics, 6, S26.

Lomolino, M.V. & Heaney, L.R. (2004) Frontiers of biogeography: new directions in the geography of nature. Sinauer Associates, Sunderland, MA.

Maddison, W.P. (1997) Gene trees in species trees. Systematic Biology, 46, 523–536.

McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T. & Glenn, T.C. (2012) Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Research, 22, 746–754.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C. & Brumfield, R.T. (2013a) Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution, 66, 526–538.

McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C. & Brumfield, R.T. (2013b) A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS ONE, 8, e54848.

Metzker, M.L. (2010) Sequencing technologies — the next generation. Nature Reviews Genetics, 11, 31–46.

Meyer, M., Briggs, A.W., Maricic, T., Hoeber, B., Hoeffner, B.H., Krause, J., Weihmann, A., Paeaebo, S. & Hofreiter, M. (2008) From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. Nucleic Acids Research, 36, e5.

Miller, T.L. & Cribb, T.H. (2007) Phylogenetic relationships of some common Indo-Pacific snappers (Perciformes: Lutjanidae) based on mitochondrial DNA sequences, with comments on the taxonomic position of the Caesioninae. Molecular Phylogenetics and Evolution, 44, 450–460.

Miller, W., Schuster, S.C., Welch, A.J., et al. (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proceedings of the National Academy of Sciences USA, 109, E2382–E2390.

Morin, P.A., Archer, F.I., Foote, A.D., et al. (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Research, 20, 908–916.

Moritz, C. & Cicero, C. (2004) DNA barcoding: Promise and pitfalls. Plos Biology, 2, e354.

O'Neill, E.M., Schwartz, R., Bullock, C.T., Williams, J.S., Shaffer, H.B., Aguilar-Miguel, X., Parra-Olea, G. & Weisrock, D.W. (2013) Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. Molecular Ecology, 22, 111–129.

Paquin, P. & Hedin, M. (2004) The power and perils of 'molecular taxonomy': a case study of eyeless and endangered Cicurina (Araneae: Dictynidae) from Texas caves. Molecular Ecology, 13, 3239–3255.

Pardo-Diaz, C., Salazar, C., Baxter, S.W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W.O. & Jiggins, C.D. (2012) Adaptive introgression across species boundaries in *Heliconius* butterflies. PLoS Genetics, 8, e1002752.

Payseur, B.A. (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. Molecular Ecology Resources, 10, 806–820.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE, 7, e37135.

Puebla, O., Bermingham, E., Guichard, F. & Whiteman, E.

 © 2013 the authors; journal compilation © 2013 The International Biogeography Society

(2007) Colour pattern as a single trait driving speciation in *Hypoplectrus* coral reef fishes? Proceedings of the Royal Society of London B, 274, 1265–1271.

Robertson, D.R., Karg, F., de Moura, R.L., Victor, B.C. & Bernardi, G. (2006) Mechanisms of speciation and faunal enrichment in Atlantic parrotfishes. Molecular Phylogenetics and Evolution, 40, 795–807.

Rocha, L.A. (2003) Patterns of distribution and processes of speciation in Brazilian reef fishes. Journal of Biogeography, 30, 1161–1171.

Rocha, L.A., Craig, M.T. & Bowen, B.W. (2007) Phylogeography and the conservation of coral reef fishes. Coral Reefs, 26, 501–512.

Rocha, L.A., Robertson, D.R., Roman, J. & Bowen, B.W. (2005a) Ecological speciation in tropical reef fishes. Proceedings of the Royal Society of London B, 272, 573–579.

Rocha, L.A., Robertson, D.R., Rocha, C.R., Van Tassell, J.L., Craig, M.T. & Bowen, B.W. (2005b) Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. Molecular Ecology, 14, 3921–3928.

Rocha, L.A., Rocha, C.R., Robertson, D.R. & Bowen, B.W. (2008a) Comparative phylogeography of Atlantic reef fishes indicates both origin and accumulation of diversity in the Caribbean. BMC Evolutionary Biology, 8, 157–157.

Rocha, L.A., Lindeman, K.C., Rocha, C.R. & Lessios, H.A. (2008b) Historical biogeography and speciation in the reef fish genus *Haemulon* (Teleostei : Haemulidae). Molecular Phylogenetics and Evolution, 48, 918–928.

Ronquist, F. & Sanmartin, I. (2011) Phylogenetic methods in biogeography. Annual Reviews of Ecology and Systematics, 42, 441–464.

Rosen, D.E. (1978) Vicariant patterns and historical explanation in biogeography. Systematic Zoology, 27, 159–188.

Rüber, L., Van Tassell, J.L. & Zardoya, R. (2003) Rapid speciation and ecological divergence in the American seven-spined gobies (Gobiidae, Gobiosomatini) inferred from a molecular phylogeny. Evolution, 57, 1584–1598.

Sanciangco, M.D., Rocha, L.A. & Carpenter, K.E. (2011) A molecular phylogeny of the Grunts (Perciformes: Haemulidae) inferred using mitochondrial and nuclear genes. Zootaxa, 2966, 37–50.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences USA, 74, 5463–5467.

Schwarz, D., Robertson, H., Feder, J., Varala, K., Hudson, M., Ragland, G., Hahn, D. & Berlocher, S. (2009) Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. BMC Genomics, 10, 633.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S. & Seeb, L.W. (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Molecular Ecology Resources, 11, 1–8.

Shendure, J. & Ji, H. (2008) Next-generation DNA sequencing. Nature Biotechnology, 26, 1135–1145.

Sites, J.W. & Marshall, J.C. (2003) Delimiting species: a renaissance issue in systematic biology. Trends in Ecology and Evolution, 18, 462–470.

Smith, K.L., Harmon, L.J., Shoo, L.P. & Melville, J. (2011a) Evidence of constrained phenotypic evolution in a cryptic species complex of agamid lizards. Evolution, 65, 976–992.

Smith, S.A., Wilson, N.G., Goetz, F.E., Feehery, C., Andrade, S.C.S., Rouse, G.W., Giribet, G. & Dunn, C.W. (2011b) Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature, 480, 364–367.

Stephen, S., Pheasant, M., Makunin, I.V. & Mattick, J.S. (2008) Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. Molecular Biology and Evolution, 25, 402–408.

Swenson, N.G. & Howard, D.J. (2005) Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. American Naturalist, 166, 581–591.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A. & Seehausen, O. (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular Ecology, 22, 787–798.

Weigt, L.A., Baldwin, C.C., Driskell, A., Smith, D.G., Ormos, A. & Reyier, E.A. (2012) Using DNA barcoding to assess Caribbean reef fish biodiversity: expanding taxonomic and geographic coverage. PLoS ONE, 7, e41059.

Woodland, D.J. (1983) Zoogeography of the Siganidae (Pisces): an interpretation of distribution and richness patterns. Bulletin of Marine Science, 33, 713–717.

Wong, E.H.K. & Hanner, R.H. (2008) DNA barcoding detects market substitution in North American seafood. Food Research International, 41, 828–837.

Wu, C.I. & Ting, C.T. (2004) Genes and speciation. Nature Reviews Genetics, 5, 114–122.

Yang, Z. & Rannala, B. (2010) Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences USA, 107, 9264–9269.

Edited by Rita Castilho