MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) **16**, 1051–1058 doi: 10.1111/1755-0998.12574

SPECIAL ISSUE: SEQUENCE CAPTURE

Capturing Darwin's dream

TRAVIS C. GLENN* and BRANT C. FAIRCLOTH†
Department of Environmental Health Science, University of Georgia, Environmental Health Science Bldg., Athens, GA 30602, USA, †Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA

Abstract

Evolutionary biologists from Darwin forward have dreamed of having data that would elucidate our understanding of evolutionary history and the diversity of life. Sequence capture is a relatively old DNA technology, but its use is growing rapidly due to advances in (i) massively parallel DNA sequencing approaches and instruments, (ii) massively parallel bait construction, (iii) methods to identify target regions and (iv) sample preparation. We give a little historical context to these developments, summarize some of the important advances reported in this special issue and point to further advances that can be made to help fulfill Darwin's dream.

Keywords: biotinylated baits, Illumina, massively parallel sequencing, next-generation sequencing, sequence capture, target enrichment

Received 14 July 2016; revision received 19 July 2016; accepted 20 July 2016

Introduction

One of Darwin's primary goals was to understand and explain 'the [evolutionary] bond' that connected the diversity of life (Darwin 1859). Although Darwin did not understand the explicit genetic mechanisms underlying shared evolutionary history, he was convinced that careful scientific study and testing of his theory would elucidate both the mechanisms and effects of evolution by natural selection. Nuttall (1901) was among the first to use molecular genetic markers, decades before DNA was shown to carry genetic information, in pursuit of Darwin's goal by studying blood protein interactions to infer evolutionary relationships. Crick, Franklin, Watson and Wilkin's discovery of the molecular structure of DNA (Franklin & Gosling 1953; Watson & Crick 1953; Wilkins *et al.* 1953) made clear the mechanism of information transfer; Sanger *et al.* (1977) gave us a method for determining the sequence of DNA; and Mullis & Faloona (1987) provided an immensely powerful tool (PCR) for selectively amplifying DNA regions of interest that could be sequenced. Combining the power of PCR with conserved primers (e.g. Kocher *et al.* 1989) or hypervariable loci (e.g. Tautz 1989) facilitated decades of research. Evolutionary insight was gained through very hard work that applied these methods to collect data from relatively few loci that were often limited to the scope of the specific questions being addressed—whether they were phylogenetic, phylogeographic, population genetic, behavioural or ecological.

The invention of massively parallel sequencing (MPS; Margulies *et al.* 2005; Bentley *et al.* 2008) fundamentally altered this *status quo* by providing a literal torrent of data while simultaneously dropping the cost of DNA sequencing to pennies per millions of bases (Glenn 2011, 2016). The ability to collect massive amounts of sequence data enabled many studies that were previously infeasible and changed a number of our assumptions about the universe of possible sequence data collection techniques (Tautz *et al.* 2010). Although MPS allows us to collect massive amounts of data at low costs, a variety of methodological, financial and analytical limitations still impede our desire to simply sequence everything (Kahvejian *et al.* 2008; Koepfli *et al.* 2015; Jones & Good 2016).

In lieu of sequencing everything, many researchers want techniques that collect data from a large number of loci across many organisms at a low, per-sample cost. Many different groups have created new ways to sample consistent, multilocus subsets of the genome from many individuals (Hardenbol *et al.* 2003; Meyer *et al.* 2007). Broadly, these methods tend to fall into one of two categories. The first uses restriction enzymes to sample a consistent portion of the genome (Baird *et al.* 2008; Elshire *et al.* 2011; Peterson *et al.* 2012), and these

Correspondence: Travis C. Glenn, Fax: 706-542-7472; E-mail: travisg@uga.edu and Brant C. Faircloth, Fax: 225-579-2597; E-mail: brant@faircloth-lab.org

techniques have taken centre stage for nonmodel organism studies at the population level (Narum *et al.* 2013). The second, which is often used in biomedical research as well as some evolutionary studies, is sequence capture (Albert *et al.* 2007; Ng *et al.* 2009; Bi *et al.* 2012; Faircloth *et al.* 2012; Lemmon *et al.* 2012; Tennessen *et al.* 2012).

Sequence capture is a type of target enrichment (Mamanova *et al.* 2010) that hybridizes single-stranded DNA or RNA baits (also called probes) to target DNA regions, physically pulls down the targeted DNA regions of interest and washes away unwanted DNA fragments so that the targeted DNA can be sequenced (Karagyozov *et al.* 1993; Kandpal *et al.* 1994; Albert *et al.* 2007; Okou *et al.* 2007; Gnirke *et al.* 2009). The background and use of sequence capture in the context of molecular ecology have recently been reviewed by Jones & Good (2016), with which we assume readers are familiar. Here, we discuss some historical context surrounding sequence capture, then we focus on the advances made by researchers with publications in this special issue, and we close with a discussion of questions and research avenues related to sequence capture that have yet to be explored.

History of sequence capture

Sequence capture is a relatively old DNA technology, with a clear history in the early 1990s, when many laboratories were developing microsatellite DNA loci (Tautz 1989; Ellegren 2004). Because microsatellites are relatively abundant in the first species investigated (~1% of cloned small DNA fragments contained microsatellite repeats in mammals), microsatellite loci could be identified by simple, though inefficient, hybridization screening of bacterial clones and Sanger sequencing of positive colonies. As the desire grew to sequence larger numbers of microsatellite loci for genome mapping purposes, and as additional workers began to focus on nonmammalian taxa where microsatellite loci are less frequent, the need increased to develop more efficient methods for microsatellite discovery and characterization.

Brenig & Brem (1991) were the first to develop a microsatellite enrichment method. Their process attached oligonucleotides composed of microsatellite repeats to a physical surface. Hybridization of DNA when one strand is attached to a surface is not efficient, so Brenig and Brem's method required large quantities of input DNA and only yielded ~10% of captured fragments with targeted microsatellites. Shortly thereafter, Ostrander *et al.* (1992) published a method that used special bacterial cloning techniques and methods derived from site-directed mutagenesis (Kunkel 1985) to enrich short-insert genomic libraries so that ~40%–50% of clones contained microsatellites. Karagyozov *et al.* (1993) and Armour *et al.* (1994) then showed that oligonucleotide probes could simply be attached to nitrocellulose filters to enrich for DNA fragments containing microsatellite repeats, but again, hybridizing DNA to probes bound to a physical surface was not efficient, although it was much easier than the microbiology required for the Ostrander *et al.* (1992) approach.

Kandpal *et al.* (1994) first demonstrated an in-solution method where they hybridized DNA to biotinylated oligonucleotides of microsatellite repeats and used beads coated with streptavidin to pull down the biotinylated probes attached to the DNA during hybridization. Kijas *et al.* (1994) then modified the approach of Kandpal *et al.* (1994) to use streptavidin-coated magnetic particles, which improved both its ease and efficiency. Further refinements of the Kijas *et al.* (1994) methods have been in widespread use for the past two decades (e.g. Hamilton *et al.* 1999; Zane *et al.* 2002; Glenn & Schable 2005), only recently being displaced by microsatellite characterization methods that simply sequence random genomic libraries using low-cost MPS (e.g. Castoe *et al.* 2012) rather than enriching, cloning, and Sanger sequencing microsatellite libraries or enrichment followed by MPS.

Shortly after MPS methods became available, the history of capturing DNA with synthetic probes was recapitulated, but this time, Albert *et al.* (2007) and others (see references in Jones & Good 2016) demonstrated that the probes could be thousands of oligonucleotides synthesized on microarrays. Gnirke *et al.* (2009) and others (see references in Jones & Good 2016) showed that cleaving the oligonucleotides from the surface of the microarray chips to facilitate in-solution hybridization of biotinylated oligonucleotides followed by pull-down with streptavidin-coated magnetic beads was superior to capturing desired targets with oligonucleotides attached to the solid surface of microarrays. In the years that have followed, nearly all companies and researchers have adopted these in-solution approaches while also adopting the term 'baits' in place of 'probes' (cf. Blumenstiel *et al.* 2010).

Special issue summary

The papers in this special issue demonstrate the varied and novel uses of in-solution sequence capture to collect genome-scale data from a large number of individuals across a variety of challenging scenarios. These difficult scenarios include data collection across different types of organisms (vertebrates, invertebrates, plants, bacteria and viruses) from different types of target regions (exons, mitochondrial DNA, ultraconserved elements [UCEs] and dsRNA) across extraordinarily different genome sizes (viruses to salamanders and pine trees) using various types of capture baits (RNA, DNA and monoclonal antibodies) from libraries made from DNA of

variable quality (from high- to low-quality modern DNA to low-quality museum DNA). Each of these papers also focuses on specific methodological advances that were needed to tackle these challenging scenarios, several of which we highlight, below.

Genome size and enrichment success

C-values of organismal genomes are the subject of much theoretical (Thomas 1971; Eddy 2012) and methodological consternation (Nystedt *et al.* 2013; Neale *et al.* 2014), and organisms having large genomes are particularly difficult to work with, even with the multitude of available techniques. Because sequence capture can explicitly target and enrich particular genomic regions, it may offer one of the few, efficient ways to collect data from organisms having large genomes. However, exactly how to make sequence capture function optimally in these organisms is poorly known. In this issue, several papers tackle this problem, demonstrating a variety of approaches to increase the efficiency of sequence capture from frogs (Portik *et al.* 2016), salamanders (McCartney-Melstad *et al.* 2016) and pines (Suren *et al.* 2016). Suren *et al.* (2016) also provide methods to deal with limitations imposed by incomplete reference genome assemblies, while Portik *et al.* (2016) and McCartney-Melstad *et al.* (2016) demonstrate the importance of different types of blocking DNA, which reduces the negative effect of non-specific hybridization, a particularly acute problem when collecting data from large-genome organisms. McCartney-Melstad *et al.* (2016) and Portik *et al.* (2016) also show the inverse relationship between input DNA concentration or library pooling and enrichment success. By contrast, Hoffberg *et al.* (2016) show that in organisms with smaller genomes and using libraries prepared with restriction enzymes, at least 96 samples can be pooled and successfully enriched for desired targets.

Sample quality

Molecular ecologists frequently deal with DNA of sub-optimal quantity and quality. These samples are sometimes collected under difficult circumstances in the field (Roffler *et al.* 2016), they may represent partially degraded DNA from gut contents (Campana *et al.* 2016), or they can be highly degraded, historical DNA extracted from museum samples (McCormack *et al.* 2016; Hawkins *et al.* 2016; Lim & Braun 2016). Even fresh samples collected from many invertebrates (Campana *et al.* 2016; Dowle *et al.* 2016; Teasdale *et al.* 2016; Yuan *et al.* 2016) and plants (Blouin *et al.* 2016; Hoffberg *et al.* 2016; Schmickl *et al.* 2016; Suren *et al.* 2016) are difficult to work with because the required extraction procedures physically damage the resulting DNA or leave impurities. Sequence capture makes these samples useable when other protocols, such as RAD-seq, may not work well with damaged or impure DNA. Lim & Braun (2016) illustrate techniques that can be used to minimize the amount and effects of damaged DNA in the resulting data.

Baits

Most papers in this issue develop new bait sets, most frequently for exons (Bragg *et al.* 2016; McCartney-Melstad *et al.* 2016; Portik *et al.* 2016; Powell *et al.* 2016; Roffler *et al.* 2016; Suren *et al.* 2016; Teasdale *et al.* 2016; Yuan *et al.* 2016), and many of the papers investigate the taxonomic range across which these newly developed baits may be useful, often focusing on the level of sequence divergence that baits can tolerate while still producing a successful outcome (Bragg *et al.* 2016; Portik *et al.* 2016; Suren *et al.* 2016). Twelve of the 15 papers in this issue use RNA baits from MYcroarray (www.mycroarray.com), whereas two studies (Bragg *et al.* 2016; Suren *et al.* 2016) use DNA baits from NimbleGen (sequencing.roche.com), one (Powell *et al.* 2016) uses RNA baits from Agilent (www.genomics.agilent.com), and one (Blouin *et al.* 2016) uses a monoclonal antibody to dsRNA. Hoffberg *et al.* (2016) take a new approach to bait design that combines the positive aspects of RAD-seq with those of sequence capture to reliably enrich polymorphic, anonymous loci from hundreds of individuals using a protocol that is exceptionally fast, easy and cost-effective.

Applications

It is possible to group the papers in this issue in a variety of ways, but regardless of grouping, it is clear that the applications of sequence capture represented by each are very diverse and include genome mapping (Suren *et al.* 2016), population genetics and phylogeography (McCormack *et al.* 2016; Hoffberg *et al.* 2016; Lim & Braun 2016; McCartney-Melstad *et al.* 2016), parasite and disease detection and ecology (Blouin *et al.* 2016; Campana *et al.* 2016; Yuan *et al.* 2016), environmental monitoring (Dowle *et al.* 2016), phylogenetics (Bragg *et al.* 2016; Hawkins *et al.* 2016; Portik *et al.* 2016; Schmickl *et al.* 2016; Teasdale *et al.* 2016) and the identification of candidate genes influenced by selection (Powell *et al.* 2016; Roffler *et al.* 2016).

Open questions and future prospects

Despite the methodological and technical advances made by these and other publications, there are a number of unanswered questions that affect how we use sequence

capture, and there are several promising avenues for future research that have not been investigated.

Not all baits are created equal

Although most researchers use in-solution biotinylated DNA or RNA baits, there are a wide variety of options for bait design and many unknowns regarding the optimality of both bait design and hybridization conditions. For example, we know little about the effects of tiling density (cf. Tewhey *et al.* 2009) or bait sequence on the downstream success of a given enrichment reaction. Similarly, we do not have a good understanding of the differences in efficiency between RNA baits (Gnirke *et al.* 2009) and DNA baits when their sequence is identical, and little empirical work focuses on understanding the relationship between bait concentration and enrichment success across organisms spanning a variety of genome sizes.

Bait length. The role of bait length relative to enrichment success is also understudied. The length of individual target enrichment baits matters, and unlike PCR primers, which are short and sensitive to 3′ mismatches, target enrichment baits are long (60–120 bases), insensitive to 3′ mismatches and tolerant to mismatches with their desired targets as a function of their length. However, there are important trade-offs regarding bait length, because longer baits (i) cost more to synthesize, (ii) will contain more synthesis errors, (iii) are limited in length by synthesis technologies, (iv) have greater potential for secondary structures and (v) take longer to hybridize. Many studies have settled on using 120-mer RNA baits to strike a balance among these factors, although in certain situations shorter baits would be a better choice for enrichments from degraded and/or formalin-fixed DNA, whereas longer baits could pull down larger fragments of DNA that are suitable for sequencing with long-read technologies (Eid *et al.* 2009; Quick *et al.* 2014; Jain *et al.* 2015).

Bait targets and phylogenetic breadth of bait conservation. Although the use of 454 sequencing and hybridization on microarrays has faded into sequence capture history, the legacy of targeting exons remains with the field. Sequence capture of exons provides researchers a variety of well-known advantages (Jones & Good 2016), and exon enrichment is the focus of nine publications in this special issue. But exons are only one of many possible genomic targets, and sequence conservation analyses clearly demonstrate an inverse relationship between exon enrichment success and phylogenetic distance (McCormack *et al.* 2012; Bragg *et al.* 2016; Jones & Good 2016; Portik *et al.* 2016). The biological reality of exon molecular evolution, where the third position of codons is more likely to vary, distributes mismatches uniformly along the lengths of divergent target regions, making exons harder to capture among divergent taxa. Simply focusing on exonic sequence may also mislead certain types of inferences due to genome-wide convergence among coding sequence (Castoe *et al.* 2009; Jarvis *et al.* 2014).

Ideally, many researchers would like to have bait sets that work across a wide range of species for a variety of purposes, so that orthologous sequence data can be collected at all levels of divergence from thousands of species. This desire has driven the development of bait sets that enrich sequence from hundreds or thousands of conserved genomic regions (Faircloth *et al.* 2012; Lemmon *et al.* 2012). Although it is logical to question the utility of enriching conserved loci when variable positions are needed, available evidence suggests that enriching and analysing these loci and the sequence that surrounds them yield variable sequences among individuals at a variety of levels of divergence (Smith *et al.* 2014; Leaché *et al.* 2015; Manthey *et al.* 2016) that may be less biased than data obtained from exons (Jarvis *et al.* 2014). That said, the effects of purifying selection at these loci (Bejerano *et al.* 2004; Harvey *et al.* 2016; Jones & Good 2016) could introduce problems for certain types of analysis, not unlike those seen for exons.

Finally, there are questions where exons and conserved loci may be too conserved to be useful (Giarla & Esselstyn 2015) or where the phylogenetic breadth of the problem is too wide—requiring thousands or tens of thousands of variable sites at the species, population and individual level. Here is where the third alternative of capturing and sequencing baits derived from variable, anonymous loci (Ali *et al.* 2016; Hoffberg *et al.* 2016) fills an important gap. First, these loci can be collected and sequenced when individual levels of resolution are needed and other marker types fail. Second, because sequence capture is such a flexible technique, it is straightforward to prepare cocktails of different bait sequences that target conserved regions, exons and anonymous loci providing the one, two, three punch of data collected at deep, moderate and (very) shallow levels simultaneously. The techniques needed to optimize this latter approach deserve additional research.

Bait synthesis. Another avenue for future research involves the process of synthesizing and resynthesizing target enrichment baits. Typically, most users order commercial baits from companies such as Agilent, MYcroarray and NimbleGen. These companies synthesize custom oligo pools with universal priming sites on each end and then use the universal primers to create RNA baits using *in vitro* transcription (Blumenstiel *et al.* 2010). The resulting pool of RNA baits is sold in limited quantity at relatively high cost. As an alternative to this full-service

approach, individual researchers could obtain custom oligo pools from MYcroarray, CustomArray (www.customarrayinc.com) or others and create their own baits (Liu *et al.* 2016) in large quantities at low cost. When smaller numbers of baits are needed, it is possible to simply order biotinylated oligonucleotides from suppliers such as IDT (www.idtdna.com), Sigma-Aldrich (www.sigmaaldrich.com) and others and use them directly for enrichment. It is also possible to use short biotinylated primers to synthesize probes as part of the enrichment process (e.g. primer extension capture, PEC; Briggs *et al.* 2009) or to use biotinylated PCR products as baits (Maricic *et al.* 2010; Peñalba *et al.* 2014). Many additional possibilities await creative minds.

Library length

In addition to the length of baits, the length of fragments in the libraries being enriched plays a critical role in sequence capture experiments. It is well appreciated that as library insert length increases, the size of the contigs that can be assembled also increases (McCormack *et al.* 2016; Jones & Good 2016). However, it may not be as well appreciated that sequencing depth (see Figure 2a of Portik *et al.* 2016) and sequence length also play critical roles. Importantly, on Illumina sequencers, the real limitation has not been read length or depth of coverage, but instead the length of fragments that can be clonally amplified for successful cluster generation and sequencing (~800 bp). As Illumina develops new technologies for cluster generation and as researchers explore other sequencing approaches, such as PacBio or Oxford Nanopore, it will be important to determine the maximum size of fragments that can be captured and sequenced successfully. Long-read technologies will likely change the game with respect to how we apply sequence capture to a given question and how we use the resulting data.

Potpourri

There are many additional areas in which sequence capture techniques could and likely will be used in the future. Growth areas are sure to include microbiome characterization, environmental DNA assessments, multilocus capture-based barcoding, host–pathogen interactions and pathogen discovery and diagnostics. To continue growth into these and other areas, it will be critical to better optimize enrichment, reduce the cost of baits and library preparations and increase the availability of baits to user groups. Experiments that investigate the effects of bait, blocker and target composition and concentration will be critical to optimize enrichment. We will need new approaches beyond dilution and sharing (see Heyduk *et al.* 2016) to reduce the cost of baits, especially when only modest numbers (dozens to hundreds) of biotinylated oligonucleotides are needed. Costs per sample also rely critically upon library preparation expenses (Meyer & Kircher 2010; Fisher *et al.* 2011; Head *et al.* 2014); thus, lower cost, highly efficient library preparation techniques that allow multiplexing of hundreds to thousands of samples per sequencing run (e.g. Glenn *et al.* 2016) will be critical.

Back to the future

Massively parallel sequencing and sequence capture are facilitating our ability to address Darwin's questions of evolutionary bonds that connect the diversity of life by enabling efficient, genome-scale data collection across an enormous number of organisms. Sequence capture techniques let us work with sample types from species having enormous genomes and those rare or extinct species with only low-quality DNA sources. We can use the data collected to study evolutionary relationships including those within families to those between species to those relating hosts and their pathogens. Sequence capture techniques also allow us to study patterns of heterozygosity, genome organization and mutation where it was not possible before. And we wonder whether the future of sequence capture will include a return to its humble roots where a single low-cost mixture of biotinylated microsatellite repeats could serve as truly universal baits. Microsatellite baits would capture both highly variable repeats and less variable flanking DNA to help understand patterns of genetic variation within any eukaryote. However, as the discovery of microsatellites has shown, sequence capture will be subsumed by our ability to simply sequence everything. Regardless of the eventual outcome, we are pretty sure Darwin would be captivated by what we can do now.

Acknowledgements

We thank the authors of each manuscript in this special issue for their contribution, A Moussalli for co-editing this special issue and A Geraldes, A DeWoody, S Narum and K Chambers for their assistance and patience in organizing and editing the special issue. Partial funding for our sequence capture work has been supported by DEB-1242260 (to BCF), DEB-1242241 (to TCG) and OISE 0730218 (to TCG) from the U.S. National Science Foundation, and P20GM103395 from the U.S. National Institutes of Health. We thank and acknowledge our many collaborators, colleagues and students who have generously assisted with our sequence capture experiments over the years. We have had a lot of fun and look forward to more in the future.

Conflict of interests

TCG declares competing interests. The EHS DNA laboratory provides oligonucleotide aliquots and services at

cost, including some services referenced in this manuscript. BCF declares no competing interests.

References

Albert T, Molla M, Muzny D *et al.* (2007) Direct selection of human genomic loci by microarray hybridization. *Nature Methods*, **4**, 903–905.

Ali OA, O'Rourke SM, Amish SJ *et al.* (2016) RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics*, **202**, 389–400.

Armour JA, Neumann R, Gobert S, Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridization selection. *Human Molecular Genetics*, **3**, 599–565.

Baird N, Etter P, Atwood T, Tirone M, Knapp SJ, Lanteri S (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Bejerano G, Pheasant M, Makunin I *et al.* (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.

Bentley DR, Balasubramanian S, Swerdlow HP *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**, 53–59.

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

Blouin AG, Ross HA, Hobson-Peters J, O'Brien C, Warren B, MacDiarmid R (2016) A new virus discovered by immunocapture of double-stranded RNA, a rapid method for virus enrichment in metagenomic studies. *Molecular Ecology Resources*, **16**, 1255–1263.

Blumenstiel B, Cibulskis K, Fisher S *et al.* (2010) Targeted exon sequencing by in-solution hybrid selection. *Current Protocols in Human Genetics/editorial board, Jonathan L. Haines... [et al.]*, **Chapter 18**, Unit–18.4.

Bragg JG, Potter S, Moritz C (2016) Exon capture phylogenomics: efficacy across scales of divergence. *Molecular Ecology Resources*, **16**, 1059–1068.

Brenig B, Brem G (1991) Direct cloning of sequence tagged microsatellite sites by DNA affinity chromatography. *Nucleic Acids Research*, **19**, 5441.

Briggs AW, Good JM, Green RE *et al.* (2009) Primer extension capture: targeted sequence retrieval from heavily degraded DNA sources. *Journal of Visualized Experiments: JoVE*, **3**, 1573.

Campana MG, Hawkins MTR, Henson LH *et al.* (2016) Simultaneous identification of host, ectoparasite and pathogen DNA via in-solution capture. *Molecular Ecology Resources*, **16**, 1224–1239.

Castoe TA, de Koning APJ, Kim H-M *et al.* (2009) Evidence for an ancient adaptive episode of convergent molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 8986–8991.

Castoe TA, Poole AW, de Koning APJ *et al.* (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS ONE*, **7**, e30953.

Darwin CR (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

Dowle EJ, Pochon X, Banks JC, Shearer K, Wood SA (2016) Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, **16**, 1240–1254.

Eddy SR (2012) The C-value paradox, junk DNA and ENCODE. *Current Biology: CB*, **22**, R898–R899.

Eid J, Fehr A, Gray J *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**, 133–138.

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews. Genetics*, **5**, 435–445.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 717–726.

Fisher S, Barry A, Abreu J *et al.* (2011) A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biology*, **12**, R1.

Franklin RE, Gosling RG (1953) Molecular configuration in sodium thymonucleate. *Nature*, **171**, 740–741.

Giarla TC, Esselstyn JA (2015) The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, **64**, 727–740.

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

Glenn TC (2016) Update to the field guide to next-generation DNA sequencers. http://www.molecularecologist.com/next-gen-fieldguide-2016. (Accessed 18 July 2016).

Glenn T, Schable N (2005) Isolating microsatellite DNA loci. *Methods in Enzymology*, **395**, 202–222.

Glenn TC, Nilsen R, Kieran TJ *et al.* (2016) Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). *bioRxiv*, 049114.

Gnirke A, Melnikov A, Maguire J *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, **27**, 182–189.

Hamilton M, Pincus E, Fiore A, Fleischer R (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques*, **27**, 500–507.

Hardenbol P, Banér J, Jain M *et al.* (2003) Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nature Biotechnology*, **21**, 673–678.

Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT (2016) Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, doi: 10.1093/sysbio/syw036.

Hawkins MTR, Hofman CA, Callicrate T *et al.* (2016) In-solution hybridization for mammalian mitogenome enrichment: pros, cons and challenges associated with multiplexing degraded DNA. *Molecular Ecology Resources*, **16**, 1173–1188.

Head SR, Komori HK, LaMere SA *et al.* (2014) Library construction for next-generation sequencing: overviews and challenges. *BioTechniques*, **56**, 61–68.

Heyduk K, Trapnell DW, Barrett CF, Leebens-Mack J (2016) Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biological Journal of the Linnean Society*, **117**, 106–120.

Hoffberg S, Kieran TJ, Catchen JM *et al.* (2016) RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, **16**, 1264–1278.

Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M (2015) Improved data analysis for the MinION nanopore sequencer. *Nature Methods*, **12**, 351–356.

Jarvis ED, Mirarab S, Aberer AJ *et al.* (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.

Jones MR, Good JM (2016) Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, **25**, 185–202.

Kahvejian A, Quackenbush J, Thompson JF (2008) What would you do if you could sequence everything? *Nature Biotechnology*, **26**, 1125–1133.

Kandpal RP, Kandpal G, Weissman SM (1994) Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 88–92.

Karagyozov L, Kalcheva ID, Chapman VM (1993) Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acids Research*, **21**, 3911–3912.

Kijas JM, Fowler JC, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *BioTechniques*, **16**(), 656–660, 662.

Kocher TD, Thomas WK, Meyer A *et al.* (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 6196–6200.

Koepfli KP, Paten B, Genome 10K Community of Scientists, O'Brien SJ (2015) The genome 10K project: A way forward. *Annual Review of Animal Biosciences*, **3**, 57–111.

Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proceedings of the National Academy of Sciences of the United States of America*, **82**, 488–492.

Leaché AD, Chavez AS, Jones LN, Grummer JA, Gottscho AD, Linkem CW (2015) Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, **7**, 706–719.

Lemmon AR, Emme SA, Lemmon EM (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–744.

Lim HC, Braun MJ (2016) High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. *Molecular Ecology Resources*, **16**, 1204–1223.

Liu S, Wang X, Xie L, *et al.* (2016) Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis. *Molecular Ecology Resources*, **16**, 470–479.

Mamanova L, Coffey AJ, Scott CE *et al.* (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods*, **7**, 111–118.

Manthey JD, Campillo LC, Burns KJ, Moyle RG (2016) Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: *Piranga*). *Systematic Biology*, **65**, 640–650.

Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Maricic T, Whitten M, Pääbo S (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE*, **5**, e14004.

McCartney-Melstad E, Mount GG, Shaffer HB (2016) Exon capture optimization in amphibians with large genomes. *Molecular Ecology Resources*, **16**, 1084–1094.

McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC (2012) Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species tree analysis. *Genome Research*, **22**, 746–754.

McCormack JE, Tsai WLE, Faircloth BC (2016) Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, **16**, 1189–1203.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**, pdb.prot5448. doi: 10.1101/pdb.prot5448.

Meyer M, Stenzel U, Myles S, Prüfer K, Hofreiter M (2007) Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Research*, **35**, e97.

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, **155**, 335–350.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

Neale DB, Wegrzyn JL, Stevens KA *et al.* (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology*, **15**, R59.

Ng SB, Turner EH, Robertson PD *et al.* (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, **461**, 272–276.

Nuttall George H F (1901) The new biological test for blood in relation to zoological classification. *Proceedings of the Royal Society of London*, **69**, 150–153.

Nystedt B, Street NR, Wetterbom A *et al.* (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature*, **497**, 579–584.

Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME (2007) Microarray-based genomic selection for high-throughput resequencing. *Nature Methods*, **4**, 907–909.

Ostrander EA, Jong PM, Rine J, Duyk G (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 3419–3423.

Peñalba JV, Smith LL, Tonione MA *et al.* (2014) Sequence capture using PCR-generated probes: a cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular Ecology Resources*, **14**, 1000–1010.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Portik DM, Smith LL, Bi K (2016) An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Molecular Ecology Resources*, **16**, 1069–1083.

Powell JH, Amish SJ, Haynes GD, Luikart G, Latch EK (2016) Candidate adaptive genes associated with lineage divergence: identifying SNPs via next-generation targeted re-sequencing in mule deer (*Odocoileus hemionus*). *Molecular Ecology Resources*, **16**, 1165–1172.

Quick J, Quinlan AR, Loman NJ (2014) A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *GigaScience*, **3**, 22.

Roffler GH, Amish SJ, Smith S *et al.* (2016) SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. *Molecular Ecology Resources*, **16**, 1147–1164.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5463–5467.

Schmickl R, Liston A, Zeisek V *et al.* (2016) Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African Oxalis (Oxalidaceae). *Molecular Ecology Resources*, **16**, 1124–1135.

Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT (2014) Target capture and massively parallel sequencing of ultraconserved elements (UCEs) for comparative studies at shallow evolutionary time scales. *Systematic Biology*, **63**, 83–95.

Suren H, Hodgins KA, Yeaman S *et al.* (2016) Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources*, **16**, 1136–1146.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, **17**, 6463–6471.

Tautz D, Ellegren H, Weigel D (2010) Next generation molecular ecology. *Molecular Ecology*, **19**(Suppl 1), 1–3.

Teasdale LC, Köhler F, Murray KD, O'Hara T, Moussalli A (2016) Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon-capture. *Molecular Ecology Resources*, **16**, 1107–1123.

Tennessen JA, Bigham AW, O'Connor TD *et al.* (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*, **337**, 64–69.

Tewhey R, Nakano M, Wang X *et al.* (2009) Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biology*, **10**, R116.

Thomas CA Jr (1971) The genetic organization of chromosomes. *Annual Review of Genetics*, **5**, 237–256.

Watson JD, Crick FHC (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.

Wilkins MHF, Stokes AR, Wilson HR (1953) Molecular structure of deoxypentose nucleic acids. *Nature*, **171**, 738–740.

Yuan H, Jiang J, Jimenez FA *et al*. (2016) Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms. *Molecular Ecology Resources*, **16**, 1095–1106.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**, 1–16.

T.C.G. and B.C.F. conceived and wrote the manuscript.