

## **Prospects for the use of Next-Generation Sequencing Methods in Ornithology**

Authors: Lerner, Heather R. L., and Fleischer, Robert C.

Source: *The Auk*, 127(1) : 4-15

Published By: American Ornithological Society

URL: <https://doi.org/10.1525/auk.2010.127.1.4>

---

BioOne Complete ([complete.BioOne.org](http://complete.BioOne.org)) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/terms-of-use](http://www.bioone.org/terms-of-use).

Usage of BioOne Complete content is strictly limited to personal, educational, and non-commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

---

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.



## SPECIAL REVIEWS IN ORNITHOLOGY

# PROSPECTS FOR THE USE OF NEXT-GENERATION SEQUENCING METHODS IN ORNITHOLOGY

HEATHER R. L. LERNER<sup>1</sup> AND ROBERT C. FLEISCHER

*Center for Conservation and Evolutionary Genetics, National Zoological Park, National Museum of Natural History,  
Smithsonian Institution, P.O. Box 37012 MRC 5503, Washington, D.C. 20013, USA*

NEARLY ALL ASPECTS of avian biology, from behavior to conservation to systematics, have benefited from the application of molecular methods over the past two decades. New technologies of high-throughput DNA sequencing are sparking a revolution in the life sciences that is sure to further transform our understanding of avian systems. Next-generation (also called “massively parallel”) sequencing methods were first introduced commercially just 5 years ago (Roche GS FLX; Margulies et al. 2005), yet their capacity to process millions of sequences in parallel, in contrast to the conventional 96-capillary capacity of Sanger sequencing, has rapidly placed them at the forefront of genetic research (Fig. 1). Although individual read lengths are currently limited (<500 base pairs [bp] for most platforms), the depth of coverage per base pair and advanced sequence-assembly software allow sequencing of 0.5–60 giga base pairs (Gbp). Thus, in a single run, depending on the platform (Table 1), it is possible to sequence anywhere from one half to 53 times the size of the chicken genome, or a minimum of 3,000 mitochondrial genomes. These methods extend far beyond simply genome sequencing and have already greatly benefited the field of biology, leading to advances in evolution (Shendure et al. 2005, Toth et al. 2007), epidemiology (Cox-Foster et al. 2007, Palacios et al. 2008), phylogenetics (Moore et al. 2006, 2007), comparative genomics (Romanov et al. 2009), microbial diversity (Edwards et al. 2006, Sogin et al. 2006, Roesch et al. 2007), DNA marker discovery, and studies of gene function and expression (Morozova and Marra 2008, Varshney et al. 2009).

These new tools make this an exciting time for ornithology because vast quantities of genetic data can be obtained affordably and with relative ease. With a little creativity, next-

generation sequencing methods (NGSMs) can be used to address new and long-standing questions previously inhibited by technological and financial limitations (e.g., obtaining enough sequence data to resolve phylogenies for adaptive radiations, study of avian major histocompatibility complex [MHC] while eliminating cloning bias, and whole-genome sequencing). At the very least, the huge increase in available DNA sequence will enable development of important markers for population genetic, systematic, behavioral, “evo-devo,” and gene mapping studies.

The benefits of NGSMs in comparison to traditional capillary-based sequencing include (1) massive amounts of sequence data for a single or multiple individuals in a single run; (2) low cost per base; (3) reduction of

the role of cloning and polymerase chain reaction (PCR) and, thus, reduced bias in resulting sequences; and (4) the ability to identify rare variant sequences rather than a single consensus sequence (see Fig. 1). There are three main drawbacks to NGSMs. The main obstacle at present, although this is beginning to change, is the relatively short read length for most platforms in comparison to traditional methods. For highly repetitive or complex genomes or genome regions, short reads (even at high coverage) may not suffice to generate an alignment. Thus, de novo sequence assembly is not always possible unless a reference or scaffold sequence of the same species or a very closely related species is available. Avian species, however,

**“These new tools make this an exciting time for ornithology because vast quantities of genetic data can be obtained affordably and with relative ease. With a little creativity, next-generation sequencing methods can be used to address new and long-standing questions previously inhibited by technological and financial limitations.”**

<sup>1</sup>E-mail: [hlerner@gmail.com](mailto:hlerner@gmail.com)

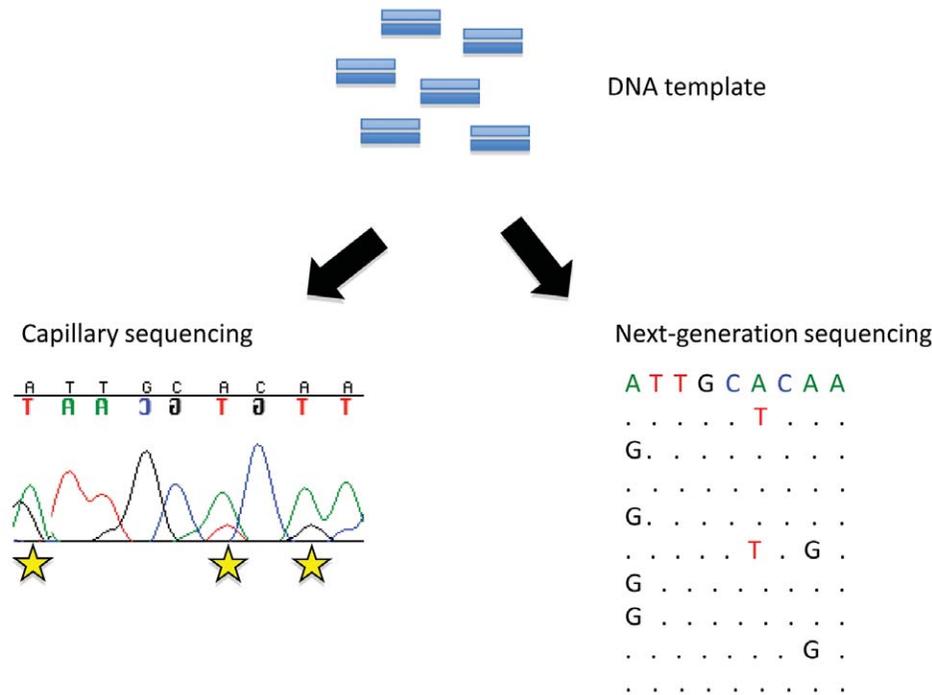


FIG. 1. Comparison of traditional capillary sequencing with next-generation sequencing methods. Capillary methods produce a single consensus read from input template in which low peaks on a chromatogram representing sequence variants cannot be distinguished from background noise. Next-generation sequencing methods produce a single sequence read for each input template, allowing the detection of low-copy variants.

are good targets for *de novo* sequence assembly, given their smaller genome sizes and lower quantity of repetitive DNA in comparison to other tetrapod taxa (Hughes and Piontkivska 2005). The benefit of generating massive amounts of data may itself pose a problem for some researchers, because bioinformatics support may be necessary to parse the data. Finally, a single run or lane may produce vastly more data than is actually necessary for a project, thus raising the cost per *required* base. Therefore, for projects that either require small amounts of sequence (in the range of a few kilobases) or have limited samples or individuals, traditional Sanger sequencing may remain the method of choice.

Clearly, the advantages of NGSs will outweigh the few drawbacks for many avian studies and study systems. In fact, certain types of questions and large data sets may only be truly approachable using NGSs. Given the relative simplicity of the laboratory techniques needed to generate next-generation sequencing libraries, we envision broad application of such methods in ornithology. Here, we provide a gateway for the avian biologist by comparing next-generation sequencing platforms, describing preparatory techniques to make such methods more broadly applicable, and discuss previous and potential applications in several subfields of avian biology.

**COMPARISON OF NEXT-GENERATION PLATFORMS**

Currently, the specifics of the chemistries and sample preparation for NGSs are largely irrelevant because the machines are too

expensive to buy or maintain for a single-researcher laboratory, with the exception of the Polonator G.007 and the recently announced 454 Junior, a next-generation bench-top pyrosequencer to be marketed in 2010 for small to medium labs. Sequencing facilities will perform much of the benchwork of sample preparation, which can be tricky depending on the platform, as part of the sequencing service for a modest fee. To choose the most appropriate platform for a project, some knowledge of the differences among methods is useful. Here, we briefly describe the methods and the unique benefits and drawbacks of each platform (see Table 1). There are several reviews that describe the technologies in greater detail, as well as animated demonstrations on the company websites (Hudson 2008, Mardis 2008, Shendure and Ji 2008; see Table 1 for company websites).

Commercial next-generation sequencing methods can be distinguished by the role of PCR in library preparation. Four main platforms are amplification-based: Roche 454 GS FLX, Illumina Genome Analyzer IIX, ABI SOLiD 3 Plus System, and Polonator G.007 (Table 1). Two single-molecule sequencing methods (i.e., not PCR-based) are either very recently available or nearly available: Helicos Genetic Analysis System and Pacific Biosciences SMRT technology, respectively (Table 1).

**Amplification-based Sequencing**

Next-generation sequencing libraries for amplification-based methods comprise short (50–400 bp) DNA templates called “sequencing features.” Preparing sequencing features requires ligation of platform-specific adapters to template DNA (Fig. 2A–C) followed by

TABLE 1. Next-generation sequencing technologies<sup>a</sup>.

Company	Roche	Applied Biosystems	Illumina	Dover Systems	Helicos	Pacific Biosciences
Website	<a href="http://www.roche-applied-science.com">www.roche-applied-science.com</a>	<a href="http://www3.appliedbiosystems.com/AB_Home/index.htm">www3.appliedbiosystems.com/AB_Home/index.htm</a>	<a href="http://www.illumina.com">www.illumina.com</a>	<a href="http://www.polonator.org">www.polonator.org</a>	<a href="http://www.helicosbio.com">www.helicosbio.com</a>	<a href="http://www.pacificbiosciences.com/">www.pacificbiosciences.com/</a>
Platform	Genome Sequencer FLX	SOLiD 3 Plus System (Supported Oligonucleotide Detection)	Genome Analyzer Iix	Polonator G.007	Helicos Genetic Analysis System	Not yet announced
Method of sequencing feature generation <sup>b</sup>	Emulsion PCR of bead-bound oligos	Emulsion PCR of bead-bound oligos	Isothermal bridge amplification on flowcell	Emulsion PCR of bead-bound oligos	NA (single-molecule sequencing)	NA (single-molecule sequencing)
Sequencing chemistry	Pyrosequencing using polymerase	Ligation (octamers, dual-base encoding)	Reversible terminator using polymerase	Ligation (nonamers, single-base encoding)	Asynchronous extension using polymerase (single-molecule synthesis)	SMART technology (single-molecule synthesis)
Availability	2005	2007	2006	2009	2008	10 test versions anticipated in early 2010, commercial version dependent on test results (anticipated in second half of 2010)
Machine cost	List price \$500,000	~\$600,000	~\$600,000	\$170,000	\$999,000	Anticipated: ~\$600,000
Reagents cost per run	~\$5,000	\$3,000 <sup>c</sup>	~\$4,000–6,000	~\$1,000	\$18,000	\$99
Capacity (base pairs) per run	0.5 Gbp	60+ Gbp, >1 billion tags per run	27.5–33 Gbp	4–5 Gbp, 8–10 Gbp with enrichment; 150 million mappable reads	28 Gbp; 400 million reads	Projected: 72 Mbp, 80,000 reads
Read length	400–500 bp (titanium chemistry) or ~250 bp (standard chemistry)	50 bp	100 bp	13 bp (26-bp paired-end reads)	25–50 bp, average 35 bp	586 bp average; 2,805 bp longest read
Run length	10 h; 2 h data processing	3.5 days	2.5–9.5 days	4 days	8 days	10–15 min minimum
Cycle length	35 s		90 min		NA	NA
Run subdivision	16	8	8 flow cells	8 lanes per flow cell, 2 flow cells	50	
Input DNA requirements	500 ng (higher for titanium chemistry)	10 ng (fragment library); 5 µg for mate-paired reads; input DNA of 200–10,000 bp in length			Application specific, as low as 50 pg	

(continued)

TABLE 1. Continued.

Company	Roche	Applied Biosystems	Illumina	Dover Systems	Helicos	Pacific Biosciences
Length of prep time	3–4 days	2–4.5 days	2 days		1 day	<12 h
Advantages	Long read length	Increased accuracy due to dual base calls, homopolymer reads sequence-able, very high output (Gbp)	No emulsion PCR	Open-source model reduces cost and promotes development	No clonal amplification, low error rate with dual pass	No clonal amplification, low cost, speed of run, direct RNA sequencing
Disadvantages	Unreliable determination of homopolymer regions and large repeats	Short read length	Medium read length, long run time	Short read length	Cost of machine	
Future developments	2010: 1 kbp read length, GSjunior bench-top platform, simplified workflow including automation of emulsion PCR and reduction in sample preparation time	100 GB and 1,200 B reads per run anticipated in February 2010	120-bp reads in early 2010, 250-bp paired-end reads in 2010	Rolling circle colonies ("rolonies") to replace emulsion PCR; 48-bp reads (24 + 24); >100 Gbp in 2010	Up to 10× improvement in throughput (no hardware change necessary) in 12–18 months; direct RNA sequencing (no cDNA intermediate) in the first half of 2010	
Platform-specific barcoding	12 barcodes	16 barcodes (with 8 plate divisions × 2 plates = 256 tags per run)	12 per flow cell			

<sup>a</sup> Information in this table was personally communicated by agents of the companies or is from other publications where noted.

<sup>b</sup> See section of text on amplification-based sequencing for explanation.

<sup>c</sup> Millar et al. 2008.

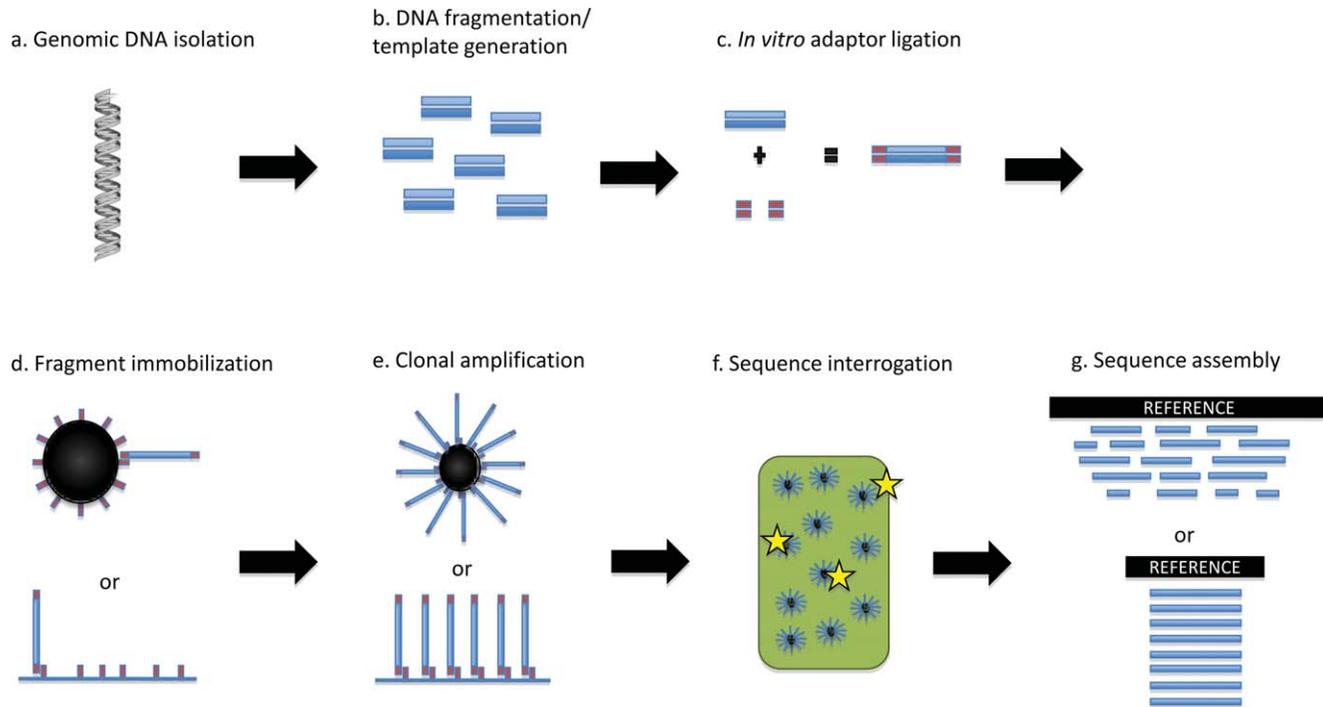


FIG. 2. A general description of the workflow common to next-generation sequencing methods (with the exception of single molecule sequencing platforms). These methods transform (a) genomic DNA to short pieces through (b) fragmentation, amplification, or DNA capture. (c) Platform-specific adaptors are then ligated *in vitro* to template sequences (barcodes for multiplexing may also be ligated at this time). (d) Using the adaptors as priming sites for hybridization, templates or fragments are immobilized to a bead (Roche 454 pyrosequencing, SOLiD and Polonator G.007) or a solid surface (Illumina GAIIx). (e) Clonal amplification is the final step in sequencing feature generation (see text section amplification-based sequencing for explanation) and is followed by (f) sequence interrogation using chemoluminescence or fluorescence. (g) Sequences are assembled to a scaffold for large genomic regions or to a reference sequence for short amplicons.

a variant of traditional PCR using the adapter sequence as priming sites or hybridization targets (Fig. 2D–E). Emulsion- or bead-based PCR (Nakano et al. 2003) is used in 454 pyrosequencing (Roche), SOLiD sequencing, and Polonator G.007 sequencing, whereas PCR employing primers covalently bonded to a flow cell (i.e., bridge PCR; Adams and Kron 1997) is used for the Illumina Genome Analyzer IIx. Following amplification, separate but parallel sequencing of each of the millions of single clonally amplified targets is performed on a substrate (Fig. 2F), for example within a micrometer-sized well on a microtiter plate (454 pyrosequencing) or directly on the tile of a flow cell (Illumina Genome Analyzer IIx).

At this point, the chemistries diverge dramatically, resulting in either large numbers of short reads (<50 bp) from Illumina, SOLiD, and the Polonator or smaller amounts of longer reads (250–500 bp) from 454 pyrosequencing. Both Illumina and Helicos use reversible dye-terminators in which a single nucleotide bound to a terminator is added by DNA polymerase and detected in real time by fluorescence, followed by removal of the terminator group (Ju et al. 2006, Mitchelson 2007). This method of “sequencing by synthesis” is continued by the addition of a different nucleotide (with terminator), and so on for a predetermined number of cycles. Pyrosequencing also uses a sequencing-by-synthesis method, but instead of reversible dye-terminators, pyrophosphate is released during nucleotide incorporation, fueling a downstream series of reactions that results in luciferase

light emission (Ronaghi et al. 1998, Margulies et al. 2005). The strength of the luciferase signal is proportional to the number of nucleotides incorporated in a template so that regions of single nucleotide repeats (homopolymers) may be read with a single light pulse. The drawback to the pyrosequencing method is that the light signal reaches an asymptote with increasing length of the single nucleotide repeat region. Thus, the ability to determine the length of a homopolymer drops off the longer the chain gets, such that homopolymers of 8 bp or longer are not reliably determined and homopolymers of as few as 4 bp can be questionable in our experience.

Alternatively, sequencing by ligation is used in SOLiD sequencing (Brenner et al. 2000, Shendure et al. 2005), where oligos of all possible di-mers (bound as degenerate 8-mers) are ligated to the single-stranded template DNA and read by fluorescence corresponding to their unique di-mers. Terminal nucleotides and the fluorescent group are cleaved, which is followed by successive rounds of ligation, detection, and cleavage up to either 25 bp or 35 bp of template DNA (depending on run specifications). The unique dual-base interrogation method of SOLiD sequencing provides very high sequencing accuracy at lower cost than many of the other methods, making it a particularly good platform for SNP discovery. The higher output (60 Gbp) from SOLiD at a lower cost per mega base pair (Mbp) also makes this an attractive sequencing method for projects that can suffice with short reads.

The Polonator G.007 also uses ligation for sequencing, but a single base rather than two bases is interrogated per ligation attempt (Dressman et al. 2003, Shendure et al. 2005). The main benefit of Polony sequencing by ligation is that the cost of the machine and recurrent reagent costs are low. Furthermore, all associated protocols and software are open-source, which maximizes flexibility.

### Single-molecule Sequencing

One single-molecule sequencing system is currently available commercially: the Helicos Genetic Analysis System, which uses technology developed by Braslavsky et al. (2003). Sequencing features are prepared for the Helicos library by the simple addition of poly-A tails to DNA templates, with no amplification of the template necessary. The template DNA is then hybridized to poly-T oligos anchored to a slide. Using a sequence-by-synthesis method (as described above) with reversible fluorescent dNTP terminators, the sequence of single-molecule templates is determined. By ligation of both a poly-A tail at the 3' end and an adaptor to the 5' end of the template, a template can be read twice, in this manner: after sequencing from the poly-A tail is completed, one strand can be removed by denaturation and the DNA can subsequently be sequenced from the 5' end using the adaptor sequence as a priming site for initiation of synthesis. This bidirectional sequencing greatly improves accuracy (from 2–7% deletion error rate with one pass to 0.2–1% with two passes, and a raw substitution rate of ~0.001% for two passes). The simplicity of library preparation and complete independence from PCR or cloning (and related errors and biases; Kanagawa 2003, Acinas et al. 2005) make this a highly attractive option (see section on ancient DNA below). Single-molecule sequencing and faster methods of massively parallel sequencing are an active area of research and development (e.g., Oxford Nanopore and Pacific Biosciences).

The speed of the five previously described NGSs is limited because each addition of a different nucleotide species (G, T, A, or C) is necessarily separated in time by several steps, particularly nucleotide detection, washing away the previous nucleotide species, and introducing a new species for incorporation. For instance, a single cycle (i.e., steps from one nucleotide incorporation to the next) requires 1–3 h, depending on the platform (with the exception of Roche 454 at 35 s cycle<sup>-1</sup>). A single molecule-sequencing method under development by Pacific Biosciences (Eid et al. 2009), SMRT observes the real-time sequencing of a single template by using a chip of thousands of nanometer-scale chambers (zero-mode waveguides; Levene et al. 2003), each with a single anchored polymerase molecule. When the DNA polymerase incorporates nucleotides with phosphate-linked fluorescent labels, a window at the bottom of the zero-mode waveguide is used to detect the fluorescence signal. As polymerase incorporates the nucleotide, it naturally cleaves the phosphate group (including the phosphate-linked fluorescent label), and through random diffusion the next nucleotide in the sequence becomes available to the polymerase. Because all four nucleotide species are present at any given time, this method is limited only by the polymerase's rate of incorporation and the machine's speed of detection. By using circularized templates and a strand-displacing enzyme, multiple independent reads can be obtained from each DNA template to

improve the sequencing accuracy to the desired level. Read lengths using the SMRT technology will also not be as limited as in other methods, currently averaging 586 bp and reaching 2,506 bp in some cases. Such long read lengths make this a particularly promising method for de novo sequencing projects. Test platforms that employ the SMRT technology will be available for preselected test laboratories in January 2010, with commercial availability dependent on their success.

## APPLICATIONS

### Multispecies, Multilocus Studies (Multiplexing)

Next-generation sequencing methods were, for the most part, originally designed with funding from genome-sequencing initiatives. Thus, runs on most platforms can be subdivided only to a limited degree (8–50 divisions; Table 1), which allows only a few individuals (or otherwise unique samples) to be sequenced in a single run. Most ecological and evolutionary studies require homologous sequence data, sometimes from multiple genomic regions and often for many individuals. High levels of sequence divergence among samples could make it possible to separate samples simply on the basis of the resulting sequence alignments (Pollock et al. 2000). For individuals of the same or a closely related species, however, samples are better distinguished by attaching a unique sequence-based tag to each sample's template before preparing the sequencing library. Both PCR-based and ligation-based techniques of tag attachment have been developed by independent researchers (Binladen et al. 2007; Meyer et al. 2007, 2008), and some additional platform-specific barcodes exist (e.g., 16 barcodes for the SOLiD 3 Plus System; Table 1). Tags can be attached either to short pieces of genomic DNA or to PCR amplicons and then pooled in equimolar ratios for preparation of the sequencing library. The pooling of tagged templates, called "multiplexing," can be done at several levels in a single library. For instance, multiple PCR amplicons of different genomic regions from the same individual can be labeled with the same tag, pooled in an equimolar ratio, and then pooled again with similar libraries that bear other unique tags. Combining tags with subdivided runs can vastly increase the number of unique samples processed in a single run to hundreds or thousands.

Targeted resequencing, also called "genome partitioning" or "DNA capture," is a revolutionary way to isolate large amounts of homologous sequence data from genomic DNA for downstream sequencing applications (Hodges et al. 2009, Summerer et al. 2009). Two primary methods exist, both based on hybridization to sequence-specific probes either in solution (Agilent and Invitrogen) or on a microarray. Such methods are particularly appropriate for studies using ancient DNA and fecal material, in which background or contaminating DNA would compete with target DNA in next-generation sequencing libraries, greatly reducing the number of useful sequences obtained (e.g., Briggs et al. 2009). At present, these methods are applicable only to single-species studies and have been applied predominantly to human studies, but development of multispecies probe applications is underway in several laboratories. Combining DNA capture with multiplexing followed by next-generation sequencing is likely to transform genetic research in the future.

### Phylogenetics

The avian tree of life at all levels stands to benefit dramatically from the orders-of-magnitude more sequence data available from NGSs. Many studies have shown that phylogenetic resolution and accuracy are improved by increasing the numbers of loci and taxa sampled (e.g., Zwickl and Hillis 2002, Rokas et al. 2003, Prasad et al. 2008). Previously unresolved avian relationships (e.g., placement of Cathartidae and Otidae; Hackett et al. 2008), particularly those that involve rapid evolutionary radiations (e.g., Hawaiian honeycreepers; Fleischer et al. 1998), may be more readily resolved with the application of NGSs to obtain massive amounts of sequence. Such “phylogenomic” analyses of rapid radiations are likely to provide insight into processes of lineage sorting, whereas studying adaptive radiations with large data sets can also provide information about functional or adaptive variation. Phylogenomics can also be applied to selected groups (e.g., ducks, warblers, and hummingbirds) to study processes of hybridization.

To date, only a few studies (all non-avian; e.g., Moore et al. 2007, Willerslev et al. 2009) have employed NGSs for phylogenetic analysis, and none has utilized the tagging methods described above. In a study that addressed the close relationships among extinct and extant rhinoceroses, Willerslev et al. (2009) used next-generation sequencing of whole mitochondrial genomes from the black, woolly, Javan, and Sumatran rhinoceroses. This data set included >16 kilobases (kbp) of mitochondrial data and provided strong support for sister relationships within the rhinoceros clade, something that studies based on smaller data sets had been unable to provide. However, higher-level relationships were not resolved with this data set, which suggests a hard polytomy in this rapid radiation that should be further investigated with data from nuclear loci.

As tagging methods become routine, next-generation sequence data of whole mitochondrial genomes and many nuclear loci are likely to become the standard in phylogenetic studies, so we may be perched on the verge of a phylogenomics era in ornithology.

### Molecular Evolution and Comparative Genomics

Patterns and processes of sequence evolution—including evolutionary rates of nucleotide and amino acid change, nucleotide and codon usage bias, and insertion-deletion patterns—can be investigated to a deeper level with the massive data sets made possible by NGSs. Next-generation sequencing methods also provide a way to investigate the unique patterns of evolution in avian-specific microchromosomes and macrochromosomes on a large scale. With the availability of multiple complete avian genomes, we anticipate new insights into the distribution of genes and recombination between the two chromosome types as well as evolution or generation of novel avian microchromosomes. Difficulties inherent to previous genome-sequencing technologies contributed to the incomplete status of the Red Jungle Fowl (i.e., chicken, *Gallus gallus*) genome. The missing sequence, predominantly on microchromosomes (5%), is being targeted with NGSs at the Genome Center at Washington University.

Comparative avian genomics, facilitated by the ease and reduced cost of genome sequencing with NGSs, has made great strides in recent years, leading to genomic resources for the

chicken, Wild Turkey (*Meleagris gallopavo*), Zebra Finch (*Taeniopygia guttata*), California Condor (*Gymnogyps californianus*), and White-throated Sparrow (*Zonotrichia albicollis*) (Romanov et al. 2009). A comparative genomic approach is extraordinarily useful for identifying functional loci related to morphological, behavioral, and physiological variation and thus enables us to better understand the process of avian evolution. For instance, sequencing multiple genomes of diverse taxa from an adaptive radiation such as the Galapagos finches or Hawaiian honeycreepers may identify the genes responsible for particular bill morphologies and other phenotypic traits. Furthermore, genome sequences for threatened taxa can be useful in developing comprehensive conservation plans that increase genetic resistance to known threats. For example, Romanov et al. (2009) performed 454 cDNA sequencing of a fibroblast cell line in California Condors and found that 78% of the reads were homologous with chicken genes, mapping to nearly all chicken chromosomes. Further analysis of this data set and additional transcriptomes is expected to identify the mutation(s) responsible for a heritable embryonic lethal condition (chondrodystrophy). Genomic data appear to be useful for managing genetic diversity in the California Condor and the goal of establishing a viable, self-sustaining population. Indeed, it is hard to overestimate the potential benefits of comparative avian genomics to species conservation and the study of avian evolution. With the advent of NGSs, avian comparative genomics will likely crescendo to the forefront of studies of avian evolution.

### Conservation and Population Genetics

Next-generation sequencing methods can contribute broadly to conservation and population genetic studies of birds (Romanov et al. 2009). Although we believe that whole-genome sequencing will be used sparingly for such studies in the near future (mainly because such studies usually require large sample sizes of individuals but do not usually have the large budgets associated with medical or agricultural applications), NGSs can still be very useful for developing variable markers such as microsatellites (Allentoft et al. 2009) and single-nucleotide polymorphisms (SNPs; e.g., Novaes et al. 2008, Vera et al. 2008) for non-model organisms. These variable markers can then be used for subsequent high-resolution analyses of population samples at lower cost.

Methods to obtain microsatellite sequences can include direct shotgun libraries or, preferably, enrichment procedures to increase the representation of microsatellite sequences within the pool of DNA fragments to be sequenced (Santana et al. 2009). The enrichment procedures followed by Santana et al. (2009) produced 873 microsatellite loci for three species (a wasp, a nematode, and a fungus) from a total of only 4.26 Mbp of sequence (14–28% of the contigs >100 bp contained a usable microsatellite locus). The amount of sequence they obtained is a bit more than can be sequenced on one 16th of a plate on a single 454 GS-FLX run or about one third of one 16th of a plate on a 454 Titanium run (at a cost of less than \$1,500 at most 454 core facilities). Allentoft et al. (2009) were able to obtain microsatellite sequences from shotgun (unenriched) 454 sequence derived from DNA extracted from bone of a late Holocene moa, *Pachyornis elephantopus*. They used the sequences to design primers for a microsatellite locus that was then successfully amplified from DNA extracted from 52 bones representing three moa genera.

Discovery and screening of SNPs generally requires that one obtain sufficient coverage (at least 5–10 times), ideally from multiple individuals (often via resequencing), to identify polymorphisms. This can be a problem for taxa with large genomes (i.e., most eukaryotes) because of the cost required to obtain full genomic sequences at that level of coverage. Thus, discovery of SNPs can be assisted by reducing the pool of homologous DNA sequences, thereby increasing the coverage level. For instance, the pool can be limited by using only transcribed sequences (the transcriptome or expressed sequence tags [ESTs]), or by restriction enzyme digestion of multiple individuals and size selection (Van Tassel et al. 2008). Vera et al. (2008) used the EST method by pooling cDNA from 80 individuals from eight families of fritillary butterflies. This method produced >600,000 EST sequences of ~110 bp in average length on the 454 GS-FLX platform. They generated 355 contigs totaling ~60,000 bp with at least 6 $\times$  coverage from the ESTs and were able to identify 751 SNPs, 149 of which involved nonsynonymous polymorphisms. Another recent study (Novaes et al. 2008) generated 148 Mbp of EST sequence from *Eucalyptus grandis* cDNA using 454 GS runs and was able to identify 23,742 high-confidence SNPs from 71,384 contigs averaging ~250 bp in length. Single-nucleotide polymorphism microarrays are available for chicken (Burt and White 2007, Schmidt et al. 2008) and Zebra Finch (Naurin et al. 2008). With >3 million SNPs identified from genomic sequences, microarray analyses with a 3-kbp SNP array from *G. gallus* or *T. guttata* can provide good coverage of the genome for association studies of diverse avian species. Given the increased length per read on the 454 with current (and near future) technology, even greater recovery of SNP sequences is expected, and we believe that application of these NGSs can greatly enhance studies in population and conservation genetics of birds.

### Disease Diagnosis and Analysis

Avian disease is of broad interest because of its potential impact on human health, its importance for conservation of rare species, and its potential for insight into the coevolution of hosts and pathogens. Because research on infectious diseases is typically well funded and involves relatively small genomes, NGSs have been applied extensively to sequencing genomes of a range of pathogens. These include viruses such as avian influenza (Höper et al. 2009) or avian bornavirus (Gancz et al. 2009), bacteria such as MRSA *Staphylococcus* (Highlander et al. 2007) and avian *Mycobacterium* (Paustian et al. 2008), and protozoans such as *Toxoplasma gondii* (Bontell et al. 2009). These methods have also been used in genome-sequencing or genetic-association studies to identify mutations responsible for genetic disorders or diseases (ten Bosch and Grody 2008, Vasta et al. 2009). In addition, researchers are beginning to apply NGSs to diagnosis or identification of pathogenic organisms by random screening (Nakamura et al. 2008, 2009; Adams et al. 2009; Jones et al. 2009) and by targeted methods using PCR amplicons (Jordan et al. 2009). Targeted, tagged pyrosequencing methods have also been used to assess variation in host immune-system genes, such as those of the MHC (Babik et al. 2009, Bentley et al. 2009, Wiseman et al. 2009) and immunoglobulins (Glanville et al. 2009). Thus far, very few NGS studies have addressed pathogen–host relationships within birds,

with the exception of a few studies on avian influenza (e.g., Höper et al. 2009) or diseases of domestic fowl (e.g., Paustian et al. 2008, Spatz and Rue 2008, Gancz et al. 2009).

### Gene Expression and Transcriptome Analyses

One of the most powerful new applications of NGSs is in measuring gene expression (Nielsen et al. 2006, Torres et al. 2007, Morozova et al. 2009). Most NGSs produce high average coverage (i.e., number of sequence copies) per nucleotide site, providing an estimation of the frequency of any particular DNA molecule in an overall DNA pool. In most expression analyses, mRNA is isolated from developing organs or from tissues that have undergone differential treatment (e.g., pathogen-infected vs. non-infected individuals), and the RNA is reverse transcribed in vitro into cDNA. With appropriate experimental controls and an assumption that the reverse-transcription process does not alter the relative frequencies of the various mRNA transcripts, the relative depth of coverage of the sequence should be proportional to the expression of the particular gene. Some early analyses (e.g., Torres et al. 2007) suggested that there were biases in the representation of certain, usually smaller, transcript sequences. But later analyses showed ways to avoid (e.g., size standardization by nebulization) or correct for these biases and revealed NGSs as a powerful tool for gene expression and evo-devo studies (Weber et al. 2007, Morozova et al. 2009). Direct RNA sequencing (i.e., without reverse transcription to cDNA) is a promising new application of the Helicos platform (Table 1) that will surely advance gene expression and transcriptome analyses.

Next-generation sequencing methods are well positioned to replace the use of more standard methods for measuring gene expression (Northern blots, real-time RT-PCR, or microarray analysis). We anticipate their use in studies of avian development such as those recently conducted by Abzhanov et al. (2004, 2006). These studies revealed, via comparative evaluations of expression levels of candidate genes in developing beak tissues and microarray analyses of expression, that the gene *Bmp4* is up-regulated in Darwin's finches with larger, thicker bills and that the calmodulin gene is up-regulated in Darwin's finches with longer, narrower bills. We also expect to see their use in studies that evaluate changes in gene expression in hosts or vectors in response to infection by avian pathogens, such as avian malaria or avian influenza, much as has been done for pathogens of agricultural or medical importance (e.g., chestnut blight; Barakat et al. 2009).

### Ancient DNA

Some of the earliest applications of next-generation sequencing involved sequencing of ancient materials such as subfossil bones, hair from mammals, and ancient soil samples (Poinar et al. 2006, Gilbert et al. 2007). Initially, these studies involved random, shotgun sequencing, but soon targeted amplification was applied to obtain full mitochondrial genomes from ancient samples (Briggs et al. 2009, Stiller et al. 2009). Most of the applications in vertebrates, thus far, have involved mammals (e.g., Noonan et al. 2005, 2006; Poinar et al. 2006; Willerslev et al. 2009), the only published use with ancient avian materials being a study on moa microsatellite development (see above; Allentoft et al. 2009).

A recent phylogenetic study (Gilbert et al. 2008) showed that sequencing of full mtDNA genomes with 454 pyrosequencing

was able to resolve two divergent clades of woolly mammoth, 1.1–1.7 million years apart, much like the pattern found for Asian elephants (Fleischer et al. 2001). Interestingly, PCR and Sanger sequencing of mtDNA control regions of 160 mammoths from Eurasia and North America suggest that this mixed pattern resulted from late Pleistocene recolonization of Eurasia by mammoth lineages that originated in North America (Debruyne et al. 2008). Two of the main obstacles in ancient DNA studies are separating the desired DNA from a high level of background DNA and removing PCR inhibitors. As described above, DNA capture methods will be useful for the former, and single-molecule sequencing could obviate the need for PCR entirely.

### Other Creative Applications

Several interesting applications of NGSMS have been developed, in addition to those discussed above. One useful method involves assessment of prey items from (noninvasive) fecal material. An exciting recent paper by Deagle et al. (2009) used PCR with redundant primers to amplify fish and cephalopod 16S (mitochondrial) and 28S (nuclear) rDNA sequences from genomic DNA isolated from Australian Fur Seal (*Arctocephalus pusillus*) feces but used blocking primers to keep the Fur Seal DNA from being amplified. The products were pyrosequenced using a 454 GS-FLX platform, and the aligned contigs were compared with databases of fish and cephalopod sequences to identify 54 species of bony fish, 4 species of cartilaginous fish, and 4 cephalopod species from 105 fecal samples. The coverage, assuming that no bias in amplification occurred, should indicate the relative frequency of the prey items in the Australian Fur Seal's diet. This study identified more diet species of Australian Fur Seals in a single analysis than were found over multiple years using traditional, extensive hard-parts analysis of samples. Comparing individuals from three distant study sites showed site-specific variation in both type and frequency of prey.

Another recent study (Soininen et al. 2009) compared the traditional DNA barcoding method and direct pyrosequencing of plant remains from stomach samples of two rodent species but found that the barcoding via PCR outperformed the pyrosequencing in identifying a range of plant taxa. We feel that a tagged, targeted approach using PCR or array capture would have greatly increased the resolution in that study. Pyrosequencing methods hold much promise for use in prey assessments for other organisms, including birds (e.g., Marrero et al. 2009).

### Data Analysis

Next-generation sequencing generates volumes of data that are both a blessing and a curse. The process of sorting quality reads and aligning and analyzing hundreds of thousands or millions of base pairs is both time intensive and computationally intensive. Technology for data acquisition is proceeding faster than information technology in many cases, and data-processing time may well exceed sample-handling time until new methods of analysis are developed. New programs and updated versions of traditional applications are being released daily to fill this gap. Given the rapidly changing nature of software programs, we have not attempted to detail them here. Several online resources feature up-to-date lists and descriptions of software packages for analysis of next-generation sequencing data. Two of the currently most useful sites are provided in the Acknowledgments.

### CONCLUSION

Next-generation sequencing methods have been applied at a blinding pace to a wide range of fields in biology and medicine. In addition, NGSMS have changed the scope and speed of standard sequencing methods by several orders of magnitude. Evolutionary biologists, ecologists, and ornithologists have been particularly slow to adopt and adapt these methods to their research programs (except perhaps a few of those involved in ancient DNA studies), but creative uses—such as targeted, tagged pyrosequencing; gene capture; and assessments of gene expression through cDNA coverage—are currently being developed. And the first whole genomes are being produced by NGSMS, with many more proposed for a huge suite of non-model organisms, including a large number and wide sampling of birds (Genome 10K Community of Scientists 2009). We predict that there will be a major acceleration in the near future in the application of NGSMS to ornithology and that many important findings in avian biology will arise from such studies. It is indeed an exciting time for ornithology.

### ACKNOWLEDGMENTS

We thank M. Meyer, M. Hofreiter, F. Hailer, D. Locke, S. Schuster, and K. Helgen for elucidation and discussion of NGSMS and the National Science Foundation (DEB-0643291) for funding. O. Ryder and A. R. Hoelzel provided valuable comments on an earlier version of the manuscript. Online resources for analysis of next-generation sequencing data are available at the following websites: [seqanswers.com/forums/showthread.php?t=43](http://seqanswers.com/forums/showthread.php?t=43), and <http://www.sanger.ac.uk/resources/software/>.

### LITERATURE CITED

- ABZHANOV, A., W. P. KUO, C. HARTMANN, B. R. GRANT, P. R. GRANT, AND C. J. TABIN. 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* 442:563–567.
- ABZHANOV, A., M. PROTAS, B. R. GRANT, P. R. GRANT, AND C. J. TABIN. 2004. *Bmp4* and morphological variation of beaks in Darwin's finches. *Science* 305:1462–1465.
- ACINAS, S. G., R. SARMA-RUPAVTARM, V. KLEPAC-CERAJ, AND M. F. POLZ. 2005. PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology* 71:8966–8969.
- ADAMS, C. P., AND S. J. KRON. 1997. Method for performing amplification of nucleic acid with two primers bound to a single solid support. U.S. Patent 5,641,658. Mosaic Technologies, Winter Hill, Massachusetts, and Whitehead Institute for Biomedical Research, Cambridge, Massachusetts.
- ADAMS, I. P., R. H. GLOVER, W. A. MONGER, R. MUMFORD, E. JACKEVICIENE, M. NAVALINSKIENE, M. SAMUITIENE, AND N. BOONHAM. 2009. Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. *Molecular Plant Pathology* 10:537–545.
- ALLENTOFT, M. E., S. C. SCHUSTER, R. N. HOLDAWAY, M. L. HALE, E. MCLAY, C. OSKAM, M. T. P. GILBERT, P. SPENCER, E. WILLERSLEV, AND M. BUNCE. 2009. Identification of microsatellites from

- an extinct moa species using high-throughput (454) sequence data. *BioTechniques* 46:195–200.
- BABIK, W., P. TABERLET, M. J. EJSMOND, AND J. RADWAN. 2009. New generation sequencers as a tool for genotyping of highly polymorphic multilocus Mhc system. *Molecular Ecology Resources* 9:713–719.
- BARAKAT, A., D. S. DILORETO, Y. ZHANG, C. SMITH, K. BAIER, W. A. POWELL, N. WHEELER, R. SEDEROFF, AND J. E. CARLSON. 2009. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BMC Plant Biology* 9:51.
- BENTLEY, G., R. HIGUCHI, B. HOGLUND, D. GOODRIDGE, D. SAYER, E. A. TRACHTENBERG, AND H. A. ERLICH. 2009. High-resolution, high-throughput HLA genotyping by next-generation sequencing. *Tissue Antigens* 74:393–403.
- BINLADEN, J., M. T. P. GILBERT, J. P. BOLLBACK, F. PANITZ, C. BENDIXEN, R. NIELSEN, AND E. WILLERSLEV. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2:197.
- BONTELL, I. L., N. HALL, K. E. ASHELFORD, J. P. DUBEY, J. P. BOYLE, J. LINDH, AND J. E. SMITH. 2009. Whole genome sequencing of a natural recombinant *Toxoplasma gondii* strain reveals chromosome sorting and local allelic variants. *Genome Biology* 10:R53.
- BRASLAVSKY, I., B. HEBERT, E. KARTALOV, AND S. R. QUAKE. 2003. Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences USA* 100:3960–3964.
- BRENNER, S., M. JOHNSON, J. BRIDGHAM, G. GOLDA, D. H. LLOYD, D. JOHNSON, S. LUO, S. MCCURDY, M. FOY, M. EWAN, AND OTHERS. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* 18:630–634.
- BRIGGS, A. W., J. M. GOOD, R. E. GREEN, J. KRAUSE, T. MARICIC, U. STENZEL, C. LALUEZA-FOX, P. RUDAN, D. BRAJKOVIĆ, Ž. KUČAN, AND OTHERS. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325:318–321.
- BURT, D. W., AND S. J. WHITE. 2007. Avian genomics in the 21st century. *Cytogenetic and Genome Research* 117:6–13.
- COX-FOSTER, D. L., S. CONLAN, E. C. HOLMES, G. PALACIOS, J. D. EVANS, N. A. MORAN, P.-L. QUAN, T. BRIESE, M. HORNIG, D. M. GEISER, AND OTHERS. 2007. A metagenomic survey of microbes in honeybee colony collapse disorder. *Science* 318:283–287.
- DEAGLE, B. E., R. KIRKWOOD, AND S. N. JARMAN. 2009. Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology* 18:2022–2038.
- DEBRUYNE, R., G. CHU, C. E. KING, K. BOS, M. KUCH, C. SCHWARZ, P. SZPAK, D. R. GRÖCKE, P. MATHEUS, G. ZAZULA, AND OTHERS. 2008. Out of America: Ancient DNA evidence for a New World origin of late Quaternary woolly mammoths. *Current Biology* 18:1320–1326.
- EDWARDS, R. A., B. RODRIGUEZ-BRITO, L. WEGLEY, M. HAYNES, M. BREITBART, D. M. PETERSON, M. O. SAAR, S. ALEXANDER, E. C. ALEXANDER, JR., AND F. ROHWER. 2006. Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* 7:57.
- EID, J., A. FEHR, J. GRAY, K. LUONG, J. LYLE, G. OTTO, P. PELUSO, D. RANK, P. BAYBAYAN, B. BETTMAN, AND OTHERS. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138.
- FLEISCHER, R. C., C. E. MCINTOSH, AND C. L. TARR. 1998. Evolution on a volcanic conveyor belt: Using phylogeographic reconstructions and K-Ar-based ages of the Hawaiian islands to estimate molecular evolutionary rates. *Molecular Ecology* 7:533–545.
- FLEISCHER, R. C., E. A. PERRY, K. MURALIDHARAN, E. E. STEVENS, AND C. M. WEMMER. 2001. Phylogeography of the Asian elephant (*Elephas maximus*) based on mitochondrial DNA. *Evolution* 55:1882–1892.
- GANCZ, A. Y., A. L. KISTLER, A. L. GRENINGER, Y. FARNOUSHI, S. MECHANI, S. PERL, A. BERKOWITZ, N. PEREZ, S. CLUBB, J. L. DERISI, AND OTHERS. 2009. Experimental induction of proventricular dilatation disease in cockatiels (*Nymphicus hollandicus*) inoculated with brain homogenates containing avian bornavirus 4. *Virology Journal* 6:100.
- GENOME 10K COMMUNITY OF SCIENTISTS. 2009. Genome 10K: A proposal to obtain whole-genome sequence for 10 000 vertebrate species. *Journal of Heredity* 100:659–674.
- GILBERT, M. T. P., J. BINLADEN, W. MILLER, C. WIUF, E. WILLERSLEV, H. POINAR, J. E. CARLSON, J. H. LEEBENS-MACK, AND S. C. SCHUSTER. 2007. Recharacterization of ancient DNA mis-coding lesions: Insights in the era of sequencing-by-synthesis. *Nucleic Acids Research* 35:1–10.
- GILBERT, M. T. P., D. I. DRAUTZ, A. M. LESK, S. Y. W. HO, J. QI, A. RATAN, C.-H. HSU, A. SHER, L. DALÉN, A. GÖTHERSTRÖM, AND OTHERS. 2008. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences USA* 105:8327–8332.
- GLANVILLE, J., W. ZHAI, J. BERKA, D. TELMAN, G. HUERTA, G. R. MEHTA, I. NI, L. MEI, P. D. SUNDAR, G. M. R. DAY, AND OTHERS. 2009. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences USA* 106:20216–20221.
- HACKETT, S. J., R. T. KIMBALL, S. REDDY, R. C. K. BOWIE, E. L. BRAUN, M. J. BRAUN, J. L. CHOJNOWSKI, W. A. COX, K.-L. HAN, J. HARSHMAN, AND OTHERS. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320:1763–1768.
- HIGHLANDER, S. K., K. G. HULTÉN, X. QIN, H. JIANG, S. YERRAPRAGADA, E. O. MASON, JR., Y. SHANG, T. M. WILLIAMS, R. M. FORTUNOV, Y. LIU, AND OTHERS. 2007. Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC Microbiology* 7:99.
- HODGES, E., M. ROOKS, Z. XUAN, A. BHATTACHARJEE, D. B. GORDON, L. BRIZUELA, W. R. MCCOMBIE, AND G. J. HANNON. 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nature Protocols* 4:960–974.
- HÖPER, D., B. HOFFMANN, AND M. BEER. 2009. Simple, sensitive, and swift sequencing of complete H5N1 avian influenza virus genomes. *Journal of Clinical Microbiology* 47:674–679.
- HUDSON, M. E. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* 8:3–17.

- HUGHES, A. L., AND H. PIONTKIVSKA. 2005. DNA repeat arrays in chicken and human genomes and the adaptive evolution of avian genome size. *BMC Evolutionary Biology* 5:12.
- JONES, R. T., R. KNIGHT, AND A. P. MARTIN. 2009. Bacterial communities of disease vectors sampled across time, space, and species. *ISME Journal*: doi:10.1038/ismej.2009.111.
- JORDAN, J. A., J. JONES-LAUGHNER, AND M. B. DURSO. 2009. Utility of pyrosequencing in identifying bacteria directly from positive blood culture bottles. *Journal of Clinical Microbiology* 47:368–372.
- JU, J., D. H. KIM, L. BI, Q. MENG, X. BAI, Z. LI, X. LI, M. S. MARMA, S. SHI, J. WU, AND OTHERS. 2006. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proceedings of the National Academy of Sciences USA* 103:19635–19640.
- KANAGAWA, T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering* 96:317–323.
- LEVENE, M. J., J. KORLACH, S. W. TURNER, M. FOQUET, H. G. CRAIGHEAD, AND W. W. WEBB. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299:682–686.
- MARDIS, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24:133–141.
- MARGULIES, M., M. EGHOLM, W. E. ALTMAN, S. ATTIYA, J. S. BADER, L. A. BEMBEN, J. BERKA, M. S. BRAVERMAN, Y.-J. CHEN, Z. CHEN, AND OTHERS. 2005. Genome sequencing in open microfabricated high-density picolitre reactors. *Nature* 437:376–380.
- MARRERO, P., R. FREGEL, V. M. CABRERA, AND M. NOGALES. 2009. Extraction of high-quality host DNA from feces and regurgitated seeds: A useful tool for vertebrate ecological studies. *Biological Research* 42:147–151.
- MEYER, M., U. STENZEL, AND M. HOFREITER. 2008. Parallel tagged sequencing on the 454 platform. *Nature Protocols* 3:267–278.
- MEYER, M., U. STENZEL, S. MYLES, K. PRÜFER, AND M. HOFREITER. 2007. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Research* 35:e97.
- MILLAR, C. D., L. HUYNEN, S. SUBRAMANIAN, E. MOHANDASAN, AND D. M. LAMBERT. 2008. New developments in ancient genomics. *Trends in Ecology and Evolution* 23:386–393.
- MITCHELSON, K. R., ED. 2007. *New High Throughput Technologies for DNA Sequencing and Genomics*. Elsevier, Amsterdam.
- MOORE, M. J., C. D. BELL, P. S. SOLTIS, AND D. E. SOLTIS. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences USA* 104:19363–19368.
- MOORE, M. J., A. DHINGRA, P. S. SOLTIS, R. SHAW, W. G. FARMERIE, K. M. FOLTA, AND D. E. SOLTIS. 2006. Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology* 6:1471–2229.
- MOROZOVA, O., M. HIRST, AND M. A. MARRA. 2009. Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics* 10:135–151.
- MOROZOVA, O., AND M. A. MARRA. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92:255–264.
- NAKAMURA, S., N. MAEDA, I. M. MIRON, M. YOH, K. IZUTSU, C. KATAOKA, T. HONDA, T. YASUNAGA, T. NAKAYA, T. HORII, AND T. IIDA. 2008. Metagenomic diagnosis of bacterial infections. *Emerging Infectious Diseases* 14:1784–1786.
- NAKAMURA, S., C.-S. YANG, N. SAKON, M. UEDA, T. TOUGAN, A. YAMASHITA, N. GOTO, K. TAKAHASHI, T. YASUNAGA, K. IKUTA, AND OTHERS. 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 4:e4219.
- NAKANO, M., J. KOMATSU, S. MATSUURA, K. TAKASHIMA, S. KATSURA, AND A. MIZUNO. 2003. Single-molecule PCR using water-in-oil emulsion. *Journal of Biotechnology* 102:117–124.
- NAURIN, S., S. BENSCH, B. HANSSON, T. JOHANSSON, D. F. CLAYTON, A.-S. ALBREKT, T. VON SCHANTZ, AND D. HASSELQUIST. 2008. A microarray for large-scale genomic and transcriptional analyses of the Zebra Finch (*Taeniopygia guttata*) and other passerines. *Molecular Ecology Resources* 8:275–281.
- NIELSEN, K. L., A. L. HOGH, AND J. EMMERSEN. 2006. DeepSAGE—Digital transcriptomics with high sensitivity, simple experimental protocol and multiplexing of samples. *Nucleic Acids Research* 34:19:e133.
- NOONAN, J. P., G. COOP, S. KUDARAVALLI, D. SMITH, J. KRAUSE, J. ALESSI, F. CHEN, D. PLATT, S. PÄÄBO, J. K. PRITCHARD, AND E. M. RUBIN. 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science* 314:1113–1118.
- NOONAN, J. P., M. HOFREITER, D. SMITH, J. R. PRIEST, N. ROHLAND, G. RABEDER, J. KRAUSE, J. C. DETTER, S. PÄÄBO, AND E. M. RUBIN. 2005. Genomic sequencing of Pleistocene cave bears. *Science* 309:597–599.
- NOVAES, E., D. R. DROST, W. G. FARMERIE, G. J. PAPPAS, JR., D. GRATAPAGLIA, R. R. SEDEROFF, AND M. KIRST. 2008. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* 9:312.
- PALACIOS, G., J. DRUCE, L. DU, T. TRAN, C. BIRCH, T. BRIESE, S. CONLAN, P.-L. QUAN, J. HUI, J. MARSHALL, AND OTHERS. 2008. A new arenavirus in a cluster of fatal transplant-associated diseases. *New England Journal of Medicine* 358:991–998.
- PAUSTIAN, M. L., X. ZHU, S. SREEVATSAN, S. ROBBE-AUSTERMAN, V. KAPUR, AND J. P. BANNANTINE. 2008. Comparative genomic analysis of *Mycobacterium avium* subspecies obtained from multiple host species. *BMC Genomics* 9:135.
- POINAR, H. N., C. SCHWARZ, J. QI, B. SHAPIRO, R. D. MACPHEE, B. BUIGUES, A. TIKHONOV, D. H. HUSON, L. P. TOMSHO, A. AUCH, M. RAMPP, W. MILLER, AND S. C. SCHUSTER. 2006. Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science* 311:392–394.
- POLLOCK, D. D., J. A. EISEN, N. A. DOGGETT, AND M. P. CUMMINGS. 2000. A case for evolutionary genomics and the comprehensive examination of sequence biodiversity. *Molecular Biology and Evolution* 17:1776–1788.
- PRASAD, A. B., M. W. ALLARD, NISC COMPARATIVE SEQUENCING PROGRAM, AND E. D. GREEN. 2008. Confirming the phylogeny of mammals by use of large comparative sequence data sets. *Molecular Biology and Evolution* 25:1795–1808.
- ROESCH, L. F. W., R. R. FULTHORPE, A. RIVA, G. CASELLA, A. K. M. HADWIN, A. D. KENT, S. H. DAROUB, F. A. O. CAMARGO, W. G. FARMERIE, AND E. W. TRIPLETT. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME Journal* 1:283–290.

- ROKAS, A., B. L. WILLIAMS, N. KING, AND S. B. CARROLL. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798–804.
- ROMANOV, M. N., E. M. TUTTLE, M. L. HOUCK, W. S. MODI, L. G. CHEMNICK, M. L. KORODY, E. M. STREMEL MORK, C. A. OTTEN, T. RENNER, K. C. JONES, AND OTHERS. 2009. The value of avian genomics to the conservation of wildlife. *BMC Genomics* 10 Supplement 2:S10.
- RONAGHI, M., M. UHLÉN, AND P. NYRÉN. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281:363–365.
- SANTANA, Q. C., M. P. A. COETZEE, E. T. STEENKAMP, O. X. MLOYNYANI, G. N. A. HAMMOND, M. J. WINGFIELD, AND B. D. WINGFIELD. 2009. Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques* 46:217–223.
- SCHMIDT, C. J., M. ROMANOV, O. RYDER, V. MAGRINI, M. HICKENBOTHAM, J. GLASSCOCK, S. MCGRATH, E. MARDIS, AND L. D. STEIN. 2008. *Gallus* GBrowse: A unified genomic database for the chicken. *Nucleic Acids Research* 36:D719–D723.
- SHENDURE, J., AND H. JI. 2008. Next-generation DNA sequencing. *Nature Biotechnology* 26:1135–1145.
- SHENDURE, J., G. J. PORRECA, N. B. REPPAS, X. LIN, J. P. MCCUTCHEON, A. M. ROSENBAUM, M. D. WANG, K. ZHANG, R. D. MITRA, AND G. M. CHURCH. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728–1732.
- SOGIN, M. L., H. G. MORRISON, J. A. HUBER, D. M. WELCH, S. M. HUSE, P. R. NEAL, J. M. ARRIETA, AND G. J. HERNDL. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proceedings of the National Academy of Sciences USA* 103:12115–12120.
- SOININEN, E. M., A. VALENTINI, E. COISSAC, C. MIQUEL, L. GIELLY, C. BROCHMANN, A. K. BRYSTING, J. H. SØNSTEBØ, R. A. IMS, N. G. YOCCOZ, AND P. TABERLET. 2009. Analysing diet of small herbivores: The efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Frontiers in Zoology* 6:16.
- SPATZ, S. J., AND C. A. RUE. 2008. Sequence determination of a mildly virulent strain (CU-2) of Gallid herpesvirus type 2 using 454 pyrosequencing. *Virus Genes* 36:479–489.
- STILLER, M., M. KNAPP, U. STENZEL, M. HOFREITER, AND M. MEYER. 2009. Direct multiplex sequencing (DMPS)—A novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Research* 19:1843–1848.
- SUMMERER, D., H. WU, B. HAASE, Y. CHENG, N. SCHRACKE, C. F. STÄHLER, M. S. CHEE, P. F. STÄHLER, AND M. BEIER. 2009. Microarray-based multicycle-enrichment of genomic subsets for targeted next-generation sequencing. *Genome Research* 19:1616–1621.
- TEN BOSCH, J. R., AND W. W. GRODY. 2008. Keeping up with the next generation: Massively parallel sequencing in clinical diagnostics. *Journal of Molecular Diagnostics* 10:484–492.
- TORRES, T. T., M. METTA, B. OTTENWÄLDER, AND C. SCHLÖTTERER. 2007. Gene expression profiling by massively parallel sequencing. *Genome Research* 18:172–177.
- TOTH, A. L., K. VARALA, T. C. NEWMAN, F. E. MIGUEZ, S. K. HUTCHISON, D. A. WILLOUGHBY, J. F. SIMONS, M. EGHOLM, J. H. HUNT, M. E. HUDSON, AND G. E. ROBINSON. 2007. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* 318:441–444.
- VAN TASSELL, C. P., T. P. L. SMITH, L. K. MATUKUMALLI, J. F. TAYLOR, R. D. SCHNABEL, C. T. LAWLEY, C. D. HAUDENSCHILD, S. S. MOORE, W. C. WARREN, AND T. S. SONSTEGARD. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* 5:247–252.
- VARSNEY, R. K., S. N. NAYAK, G. D. MAY, AND S. A. JACKSON. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27:522–530.
- VASTA, V., S. B. NG, E. H. TURNER, J. SHENDURE, AND S. H. HAHN. 2009. Next generation sequence analysis for mitochondrial disorders. *Genome Medicine* 1:100.
- VERA, J. C., C. W. WHEAT, H. W. FESCEMYER, M. J. FRILANDER, D. L. CRAWFORD, I. HANSKI, AND J. H. MARDEN. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* 17:1636–1647.
- WEBER, A. P., K. L. WEBER, K. CARR, C. WILKERSON, AND J. B. OHLROGGE. 2007. Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiology* 144:32–42.
- WILLERSLEV, E., M. T. P. GILBERT, J. BINLADEN, S. Y. W. HO, P. F. CAMPOS, A. RATAN, L. P. TOMSHO, R. R. DA FONSECA, A. SHER, T. V. KUZNETSOVA, AND OTHERS. 2009. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evolutionary Biology* 9:95.
- WISEMAN, R. W., J. A. KARL, B. N. BIMBER, C. E. O’LEARY, S. M. LANK, J. J. TUSCHER, A. M. DETMER, P. BOUFFARD, N. LEVENKOVA, C. L. TURCOTTE, AND OTHERS. 2009. Major histocompatibility complex genotyping with massively parallel pyrosequencing. *Nature Medicine* 15:1322–1326.
- ZWICKL, D. J., AND D. M. HILLIS. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51:588–598.

Received 30 November 2009, accepted 12 December 2009.