Improving Genome Assemblies and Annotations for Nonhuman Primates

Robert B. Norgren Jr.

Abstract

The study of nonhuman primates (NHP) is key to understanding human evolution, in addition to being an important model for biomedical research. NHPs are especially important for translational medicine. There are now exciting opportunities to greatly increase the utility of these models by incorporating Next Generation (NextGen) sequencing into study design. Unfortunately, the draft status of nonhuman genomes greatly constrains what can currently be accomplished with available technology. Although all genomes contain errors, draft assemblies and annotations contain so many mistakes that they make currently available nonhuman primate genomes misleading to investigators conducting evolutionary studies; and these genomes are of insufficient quality to serve as references for NextGen studies. Fortunately, NextGen sequencing can be used in the production of greatly improved genomes. Existing Sanger sequences can be supplemented with NextGen whole genome, and exomic genomic sequences to create new, more complete and correct assemblies. Additional physical mapping, and an incorporation of information about gene structure, can be used to improve assignment of scaffolds to chromosomes. In addition, mRNA-sequence data can be used to economically acquire transcriptome information, which can be used for annotation. Some highly polymorphic and complex regions, for example MHC class I and immunoglobulin loci, will require extra effort to properly assemble and annotate. However, for the vast majority of genes, a modest investment in money, and a somewhat greater investment in time, can greatly improve assemblies and annotations sufficient to produce true, reference grade nonhuman primate genomes. Such resources can reasonably be expected to transform nonhuman primate research.

Key Words: ape, evolution; genome annotation; genome assembly; lemur; monkey; nonhuman primates; translational research

Dr. Robert B. Norgren, Jr., Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, Nebraska.

Address correspondence and reprint requests to Dr. Robert B. Norgren, Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, 985805 Nebraska Medical Center, Omaha, NE 68198 or email rnorgren@unmc.edu.

Introduction

Nonhuman primates include the great apes—chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), gorillas (*Gorilla gorilla*), and orangutans (*Pongo abelii*); the lesser apes including the gibbon (*Nomascus leucogenys*); old world monkeys including rhesus macaques (*Macaca mulatta*), cynomolgus monkeys (*Macaca fascicularis*), pigtail macaques (*Macaca nemestrina*), baboons (*Papio anubis*), African green monkeys (genus: *Chlorocebus*), and sooty mangabeys (*Cercocebus atys*); new world monkeys including marmosets (*Callithrix jacchus*); and strepsirrhines including the aye-aye (*Daubentonia madagascariensis*), the grey mouse lemur (*Microcebus murinus*), and the sifaka (*Propithecus coquerelli*). Draft genomes for all of the species listed above are either in progress or are complete (Table 1; Marques-Bonet et al. 2009). This vast group of species represents a significant span of primate evolutionary history (Figure 1). These animals are used in two main types of studies: the study of human evolution and as models for translational medicine. Both types of study would benefit from high quality genomic information, a resource that is currently not available for any nonhuman primate.

Evolutionary studies in nonhuman primates are important, not only for their intrinsic intellectual interest, but also because filtering genetic variants identified in NextGen sequencing studies, designed to determine which mutations are related to disease, requires a knowledge of evolutionary context (MacArthur et al. 2012; Torkamani et al. 2012; Norgren 2012). High quality genomes are important for these studies because the comparison of genes across species cannot yield correct results if the genes for a given species have not been annotated, or have been incorrectly annotated (Vallender 2009). To conduct these studies, genomic information from species closely related to humans like chimpanzees and species more distantly related like the lesser apes, and monkeys, are necessary for correct results.

Biomedical researchers engaged in translational research depend on nonhuman primates because these animals are most likely to recapitulate human symptoms, and offer the best hope for predicting human responses to experimental therapies. Rhesus macaques are the most frequently used animal models among the nonhuman primates. This species is used extensively for investigations into the pathogenesis of AIDS, and countermeasures including vaccines (Baroncelli

Table 1 Status of nonhuman primate genome sequencing projects

Common name	Scientific name	Coverage				Status	Refs
		Sanger	Illumina	454	Solid PacBio		
Rhesus macaque	*Macaca mulatta*	6X				Complete	1
Rhesus macaque	*Macaca mulatta*		47X			Complete	2
Cynomolgus macaque	*Macaca fascicularis*		54X			Complete	2
Cynomolgus macaque	*Macaca fascicularis*				40X	Complete	3
Cynomolgus macaque	*Macaca fascicularis*			3X	3X	Complete	4
Chimpanzee	*Pan troglodytes*	6X				Complete	5
Chimpanzee	*Pan troglodytes verus* (10 samples)		9X			Complete	6
Chimpanzee	P*an troglodytes* (25 samples, 4 subspecies)		23X			Complete	7
Bonobo	*Pan paniscus*			26X		Complete	8
Bonobo	*Pan paniscus* (13 samples)		27X			Complete	7
Gorilla	*Gorilla gorilla*	2X	60X			Complete	9
Gorilla	*Gorilla gorilla* (27 samples, 3 subspecies)		18X			Complete	7
Orangutan	*Pongo abelii*	6X				Complete	10
Orangutan	*Pongo abelii/pygmaeus* (10 samples)		9.3X			Complete	10
Orangutan	*Pongo abelii/pygmaeus* (10 samples)		27X			Complete	7
Aye-aye	*Daubentonia madagascariensis*		38X			Complete	11
Common marmoset	*Callithrix jacchus*	6X				In progress	12
White-cheeked gibbon	*Nomascus leucogenys*	6X				In progress	13
Olive baboon	*Papio anubis*	2X	85X	4.5X		In progress	14
African green monkey	*Chlorocebus aethiops*			18X		In progress	15
Mouse lemur	*Microcebus murinus*	2X	100X		7X	In progress	16, 17
Sooty mangabey	*Cercocebus atys*		107X		6X	In progress	17
Pigtail macaque	*Macaca nemestrina*		100X		6X	In progress	17
Sifaka	*Propithecus coquerelli*		100X		6X	In progress	17

Note: Sequences from different platforms but from the same individual are on the same horizontal line. Sequences from the same species but different individuals are indicated on different horizontal lines. When information is available on the likely platform and amount of coverage for "in progress" species is available, estimates are provided.

References: 1. Gibbs et al. (2007); 2. Yan et al. (2011); 3. Higashino et al. (2012); 4. Ebeling et al. (2011); 5. Chimpanzee Sequencing and Analysis Consortium (2005); 6. Auton et al.; 7. Prado-Martinez et al. (2013); 8. Prüfer. et al. (2012); 9. Scally et al. (2012); 10. Locke et al. (2011); 11 Perry et al. (2012); 12. http://genome.wustl.edu/genomes/detail/callithrix-jacchus/ (accessed on July 19, 2013); 13. http://genome.wustl.edu/genomes/detail/nomascus-leucogenys/ (accessed on July 19, 2013); 14. http://www.ncbi.nlm.nih.gov/nuccore/AHZZ00000000.1/ (accessed on July 19, 2013); 15. http://genome.wustl.edu/genomes/detail/chlorocebus-aethiops/ (accessed on July 19, 2013); 16. https://www.hgsc.bcm.edu/content/mouse-lemur-genome-project (accessed on July 19, 2013); 17. Personal communication, Dr. Jeff Rogers.

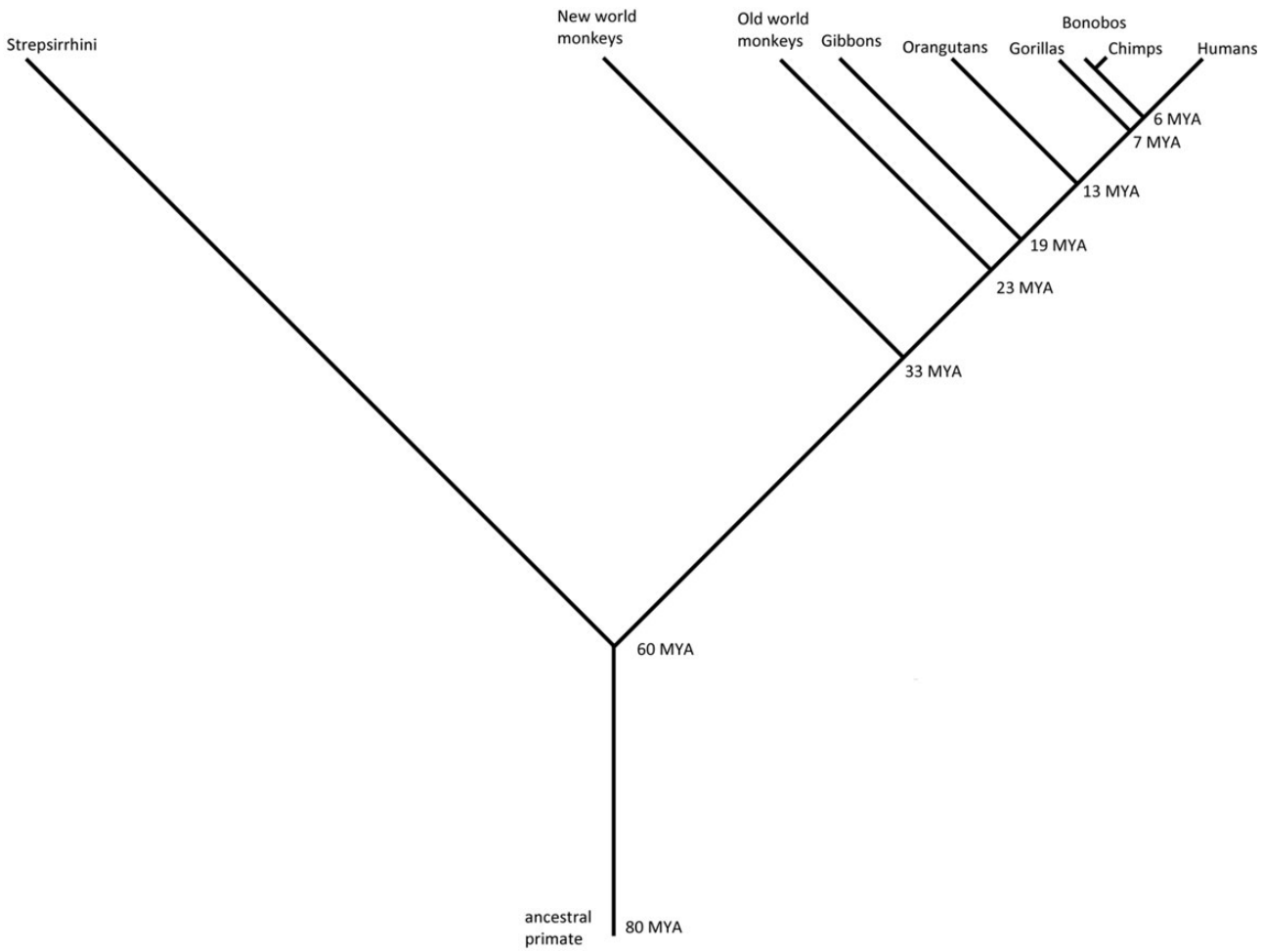

Figure 1: Phylogeny of nonhuman primates The point of divergence from human's the last common ancestor is indicated at the branching points in millions of years ago (MYA). Strepsirrhines include the aye-aye (*Daubentonia madagascariensis*), grey mouse lemur (*Microcebus murinus),* and sifaka (*Propithecus coquerelli*). New world monkeys include the common marmoset (*Callithrix jacchus*). Old world monkeys include rhesus macaques (*Macaca mulatta*), cynomolgus monkeys (*Macaca fascicularis*), pigtail macaques (*Macaca nemestrina*), baboons (*Papio anubis*), African green monkeys (genus: *Chlorocebus*), and sooty mangabeys (*Cercocebus atys*). Gibbons are lesser apes, including the white-cheeked gibbon (*Nomascus leucogenys*). Orangutans (*Pongo abelii*), gorillas (*Gorilla gorilla*), chimpanzees (*Pan troglodytes*), and bonobos (*Pan paniscus*) are all great apes. Bonobos and chimpanzees last shared a common ancestor about one million years ago. Draft genomes are (or will be) available for all the species listed (see Table 1).

et al. 2008; Shedlock et al. 2009). Sooty mangabeys and African green monkeys are also of interest to AIDS investigators because these species are naturally resistant to HIV (Bosinger et al. 2011; Chahroudi et al. 2012; Ma et al. 2013). Pharmaceutical companies use cynomolgus monkeys extensively to test drugs before advancing to clinical trials because the toxicological profile of a drug can be different in rodents than in humans and nonhuman primates (Chellman et al. 2009). Nonhuman primates are also used for studies of higher order cognitive function and behavior (Watson and Platt 2012; Fox et al. 2012; Shackman et al. 2013), reproduction (Hewitson 2004; Sparman et al. 2007), and cardiovascular disease (Vinson et al. 2011). All of these studies could benefit from NextGen sequencing approaches, especially mRNA-seq for expression analysis, and exome sequencing for investigations of genetic effects.

Most NextGen sequencing studies require a reference genome sequence against which to align the short sequences in these studies. The current draft rhesus genome has been found to be inadequate as a reference genome. For example, investigators conducting expression studies in rhesus macaques have not been able to obtain high quality data when aligning rhesus NextGen mRNA sequences against the draft genome (Dr. J Knowles, University of Southern California, Keck School of Medicine, personal communication, 2013, and Dr. N Kalin, University of Wisconsin School of Medicine and Public Health, personal communication, 2013). Attempts to use the rhesus genome for NextGen Exome studies have also revealed serious problems (Vallender 2011). Similar issues likely apply to the genomes for other nonhuman primates.

Limitations of Draft Nonhuman Primate Genomes and Annotations

The "draft" human genome was published with 7.5X Sanger coverage (Lander et al. 2001). A considerable amount of additional "finishing" Sanger sequencing was performed to produce a higher quality human genome assembly (International Human Genome Sequencing Consortium, 2004). Additional "finishing" of human chromosomes continued to 2006 (Dhand 2006). In contrast, nonhuman primates have been sequenced with, at most, 6X Sanger coverage (Table 1). Three species have received this level of Sanger coverage: rhesus macaques (Gibbs et al. 2007), orangutans (Locke et al. 2011), and chimpanzees (Chimpanzee Sequencing and Analysis Consortium 2005). It is anticipated that a similar level of coverage will be obtained by Sanger sequencing for the common marmoset (http://genome.wustl.edu/genomes/detail/callithrix-jacchus/) and white-cheeked gibbons (http://genome.wustl.edu/genomes/detail/nomascus-leucogenys/). The gorilla (Scally et al. 2012), bonobo (Prüfer. et al. 2012), and cynomologus macaque (Yan et al. 2011) were sequenced with a mixture of Sanger and NextGen methods, or NextGen methods alone (Table 1). As a result of shorter read lengths, many more NextGen sequences are needed to produce Sanger level coverage. This fact makes comparisons of coverage between the different technologies difficult. The mixed and NextGen only genomes are likely to be even less complete than the Sanger only genomes. Thus, the available nonhuman primate genomes have received far less attention to quality of assembly than the human genome.

Since the quality of the rhesus macaque genome has been independently assessed by a number of methods and it was finished to a similar level as the other nonhuman primate genomes, this species will be used as an example of the limitations of draft nonhuman primate genomes. Radiation hybrid (Karere et al. 2008) and FISH analyses (Roberto et al. 2008) have demonstrated numerous misassemblies in the rhesus genome. Using a gene-based approach (Zhang et al 2012), several types of misassemblies have been documented including: scaffolds assigned to the wrong chromosome (Figure 2A), and scaffolds placed in the wrong orientation (Figure 2B). In addition, there were many scaffolds containing entire genes or fragments of genes that were not placed on any chromosome (Figure 2B). There are documented sequencing errors that introduced apparent nonsense mutations in exons (Figure 2C). The assembly and sequencing errors cause automated annotators to incorrectly or fail to annotate many genes. Further, missing exons can result in automated annotation pipelines, which generate spurious sequences. Automated annotators may use intronic sequence to create incorrect gene models (Figure 2D). Serious errors in annotation of the rhesus genome have also been identified in the course of attempting to use it for evolutionary studies and NextGen exome analysis (Vallender 2009, 2011). It is estimated that approximately 50% of the rhesus gene annotations available at NCBI are either incomplete or incorrect (Zhang et al. 2012). It is important to note that this is not just an issue for nonhuman primate genomes. Similar problems have been observed for other draft genomes as well (Nagy 2008, 2011).

Assemblies have been reported for a Chinese rhesus macaque (Yan et al. 2011) and several cynomolgus macaques (Ebeling et al. 2011; Yan et al. 2011; Higashino et al. 2012). However, because these assemblies were all dependent on the original chromosome files produced for the reference rhesus animal, they can be expected to contain many of the same misassemblies documented for the rhesus genome.

Comparing draft nonhuman primate genomes with the human genome will become increasingly difficult as NextGen sequencing is either combined with existing Sanger sequences or used separately. There are no clear standards or benchmarks for assessing the quality of draft sequences. However, there have been independent efforts at evaluating different assemblers (Narzisi and Mishra 2011; Earl et al. 2011; Salzberg et al. 2012). The N50 statistic is often reported for assemblies. The bigger the number, the longer the contigs contained within an assembly. Although it may be tempting to judge reports of bigger N50s as better, this is not necessarily the case (Narzisi and Mishra 2011; Salzberg et al. 2012). Scaffolds, and even contigs, can be misjoined, that is, sequences are placed together that should be far apart, perhaps even on different chromosomes (Salzberg et al. 2012; Zhang et al. 2012). Depending on the assembler chosen, one can get smaller N50s with more accurate assemblies, or bigger N50s with more mistakes in the assembly (Narzisi and Mishra 2011; Salzberg et al. 2012). In the final analysis, the most important quality metric for most users is the number of genes that can be correctly and completely annotated for a given genome. Unfortunately, annotation reports in most mammalian genome papers are so terse as to make it difficult to assess their completeness or correctness.

Strategies for Improving Nonhuman Primate Genomes

Finishing a draft genome, to the level of the human genome, using the same approaches would be prohibitively expensive. There is considerable interest in using much less expensive NextGen sequencing to "top off" Sanger sequencing, or replace it entirely. At this time, there is no standardized approach for using NextGen sequences to produce a high quality genome. However, for animals for where there are already Sanger sequences available, it makes sense to utilize a hybrid strategy that combines the long reads from the Sanger sequencing with NextGen Sequences (Figure 3). Ideally, the genomic DNA used for the NextGen sequencing should come from the same animal that was used for the Sanger sequencing.

After contigs and scaffolds are created using an assembly pipeline, they must be examined for misassemblies. It is not uncommon for scaffolds to contain fragments that belong in different locations in the same chromosome, or even on

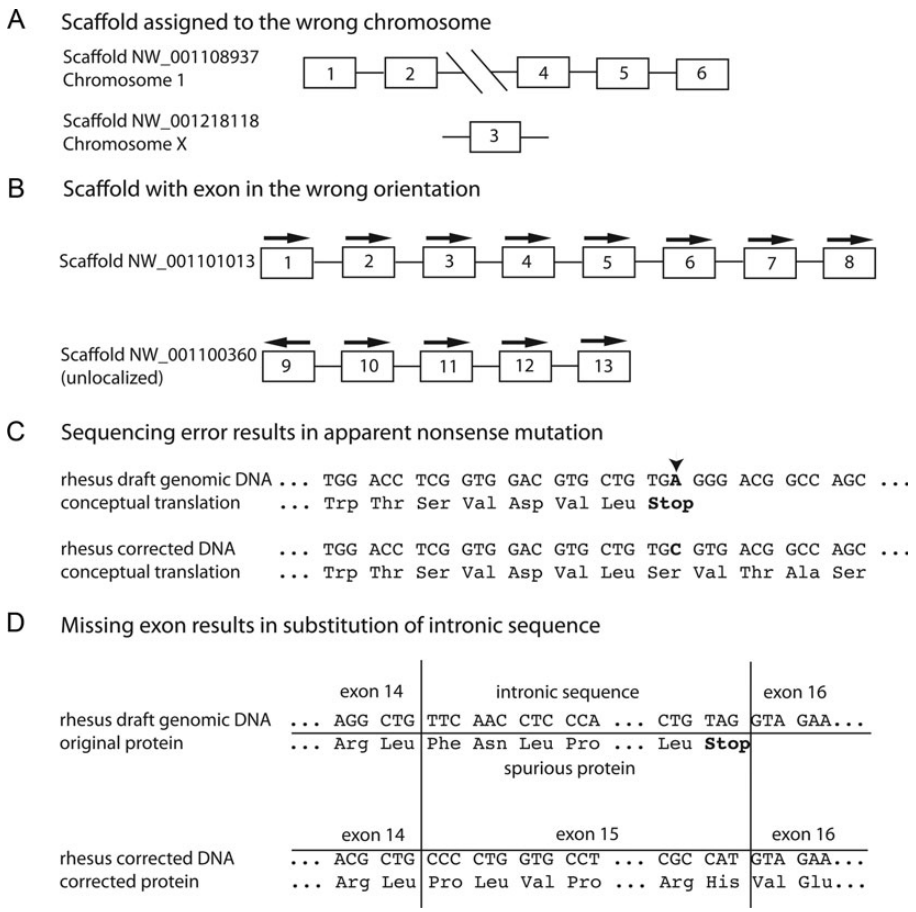

Figure 2: Schematic diagrams illustrating assembly and annotation errors in the rhesus macaque draft genome. (A) Scaffold assigned to the wrong chromosome: The scaffold containing exons 1, 2, 4, 5, and 6 of the SRC homology 2 domain containing E (SHE) gene is correctly assigned to chromosome 1 in the rhesus draft genome. However, the scaffold containing exon 3 of the SHE gene was incorrectly assigned to chromosome X. (B) Scaffold with exon in the wrong orientation: An unlocalized scaffold from the draft rhesus genome contains exons 9-13 of the Bardet-Biedl syndrome 1 (BBS1) gene. It was not included in the rhesus chromosome 14 file with the scaffold that contains exons 1-8 of BBS1. This is likely the contig containing exon 9 was in the wrong orientation with respect to the rest of the scaffold. (C) Sequencing error results in apparent nonsense mutation: The rhesus draft genomic DNA had sequencing error in the adrenergic, beta-1, receptor (ADBR1) gene. This introduced a premature stop codon (arrow, top panel). This has resulted in this locus being labeled a pseudogene by NCBI. Our targeted sequencing of this region has revealed the correct sequence (JN589014.1 - bottom panel). (D) Missing exon results in substitution of intronic sequence: The original rhesus draft genome did not contain the sequence for exon 15 for the adenylate cyclase 3 (ADCY3) gene. Instead, intronic sequence between exons 14 and exon 16 was substituted (top panel) when this gene was annotated. This led to spurious protein sequence (original protein) and a premature stop codon. The missing exon 15 was sequenced and deposited in GenBank (HM067826.1). NCBI then corrected the rhesus ADCY3 gene model and now reports a correct protein sequence for this gene (bottom panel). Figure 2 is redrawn from Zhang et al. 2012.

different chromosomes (Figure 2A). This is because repeat regions are scattered throughout the genome and can confuse assemblers. One test relating to the correctness of scaffolds, examines exon order within genes. Since exon order is completely conserved among mammals, exons for a given gene should be in the same order and orientation in all mammals. This information can be used not only to identify misassembled scaffolds, and correct them, but also to place scaffolds in the correct order. If two different scaffolds each contain some exons from the same gene, then they must belong adjacent to each other.

Synteny is relatively well conserved among the apes and old world monkeys. Still, there have been a number of chromosome fission, chromosome fusion, and rearrangements in the course of primate evolution (Kehrer-Sawatzki and Cooper 2008; Capozzi et al. 2012; Nie 2012; de Oliveira et al. 2012; Stanyon et al. 2012). Thus, to place scaffolds on chromosomes accurately, mapping information is necessary. Three types of maps are available for the rhesus macaque: genetic (Rogers et al. 2006), radiation hybrid (Karere et al. 2008), and FISH (Roberto et al. 2008). Although helpful, even more detailed maps would be useful when placing scaffolds correctly on chromosomes. For other nonhuman primates, the amount of mapping data varies, but is generally less than ideal.

Improved assemblies are necessary to make nonhuman primate genomes useful, but they are not sufficient.

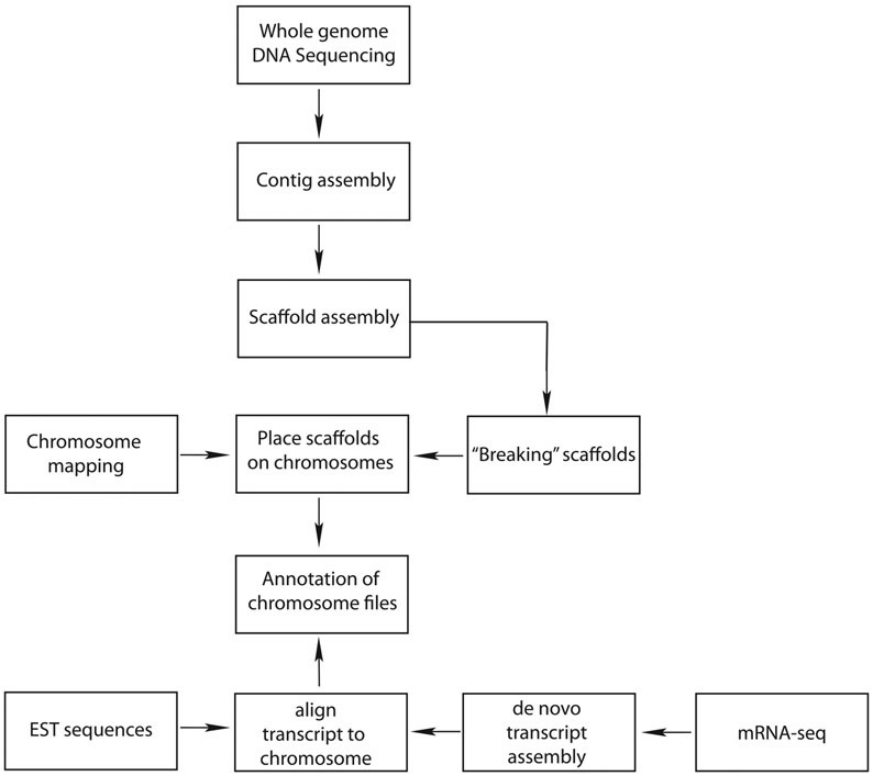

Figure 3: Flowchart describing assembly and annotation procedures The steps involved in creating a high-quality genome. Sequencing can include the conventional Sanger technique and/or several NextGen technologies including 454, Illumina, and Ion Torrent (see Table 1). Contig and scaffold assembly can utilize several assemblers including: Atlas (Havlak et al. 2004), AbySS (Simpson et al. 2009), ALLPATHS-LG (Gnerre et al. 2011), Celera assembler (Myers et al. 2000), MaSuRCA (http://www.genome.umd.edu/masurca.html) (accessed on July 19, 2013), and SOAPdenovo (Li et al. 2010). Chromosome mapping can use genetic information, radiation hybrids or fluorescence *in situ* hybridization (FISH). "Breaking" misassembled scaffolds and placing them on chromosomes can involve extensive manual work. Expressed sequence tags (ESTs) are usually partial transcripts obtained from Sanger sequencing. mRNA-seq is often performed with Illumina technology but can also be conducted with Ion Torrent machines.

Annotations, specifications of exon ranges from named genes within chromosome sequences, are required before nonhuman primate genomes can be used for NextGen expression and exome studies. Many regions of draft nonhuman primate genomes have provisional annotations. For example, a region may be annotated with a "gene" which is designated with "LOC" prefixed to a set of numbers. Although this may be counted as an annotation, it is of little value to users. Investigators prefer that orthologs of human genes be annotated with the same names so that comparisons between human and nonhuman primate data can be easily made.

Annotations of nonhuman primate assemblies can be improved by aligning named transcripts against the chromosome files of the genome to be annotated using programs like sim4cc (Zhou et al. 2009), GMAP (Wu and Watanabe 2005), and Spidey (Wheelan et al. 2001). Due to their close evolutionary relationships, the well-characterized human RefSeq transcripts can be used to identify orthologs in most primate species. However, mRNA transcripts from the closest species to the target genome will work the best. Ideally, one would use transcripts from the same species being annotated. Some expressed sequence tags (ESTs) are available for nonhuman primates (Magness et al. 2005). However, in no case are these complete. One way to economically and efficiently obtain transcripts is to perform NextGen, mRNA sequencing from cells and tissues derived from the species of interest. (Table 2; Pipes et al. 2013; Perry et al. 2012) Transcripts can then be constructed using *de novo* assemblers such as velvet or oases (Zerbino and Birney 2008; Schulz et al. 2012), and Trinity (Grabherr et al. 2011).

For the vast majority of nonhuman primate transcripts, establishing orthology with human genes is straightforward. However, for highly divergent, and highly polymorphic genes, such as those contained within the MHC, annotation is much more difficult. For these regions, the first challenge is the assembly itself. It is difficult to get a good assembly of the MHC region without careful bacterial artificial chromosome (BAC) sequencing, as has been performed for the rhesus macaque (Daza-Vamenta et al. 2004). Long, accurate reads are currently not possible with NextGen sequencing, but may be available in the near future. Although this will aid assembly of the MHC region, annotation will still be a challenge. Because there is so much variation in MHC genes

Table 2 Nonhuman primate transcriptome projects

Common name	Scientific name	Refs
Chimpanzee	*Pan troglodytes*	1, 2
Gorilla	*Gorilla gorilla*	2
Rhesus macaque	*Macaca mulatta*	1, 2
Cynomolgus macaque	*Macaca fascicularis*	2
Japanese macaque	*Macaca fuscata*	2
Pig-tailed macaque	*Macaca nemestrina*	2
Olive baboon	*Papio anubis*	2
African green monkey	*Chlorocebus aethiops*	1, 2
Sooty mangabey	*Cercocebus atys*	2
Common marmoset	*Callithrix jacchus*	1, 2
Owl monkey		2
Squirrel monkey		2
Mohol bushbaby	*Galago moholi*	1
Slow loris		1
Aye-aye	*Daubentonia*	1
Madagascariensis		
Coquerel's sifaka	*Propithecus coquereli*	1
Black & white ruffed lemur	*Varecia variegata*	1
Ring-tailed lemur	*Lemur catta*	2
Mongoose lemur	*Eulemur mongoz*	1
Crowned lemur	*Eulemur coronatus*	1
Mouse lemur	*Microcebus murinus*	2

References: 1. Perry et al. (2012); 2. Pipes et al. (2013)

among individuals of any nonhuman primate species, no one annotation of an individual's genome will encompass all of the MHC genes. Extensive characterization of MHC transcripts followed by manual annotation of chromosome files will be required. For highly polymorphic immune system genes, specialized databases such as the International ImMunoGeneTics information system (http://www.imgt.org/) will be useful adjuncts to the reference genome.

Future Opportunities

High quality nonhuman primate genomes have the potential to catalyze biomedical research. A rhesus macaque GeneChip, in collaboration with Affymetrix has been developed (Spindel et al. 2005; Duan et al. 2007). Although this has proven to be a useful reagent for gene expression studies for rhesus macaques, mRNA-seq using NextGen sequencing promises cheaper and more sensitive expression studies. Further, expression microarrays are not available for most nonhuman primates. State-of-the-art expression studies can be performed in any nonhuman primate species once a high quality genome is available. These studies will likely accelerate biomedical research.

Exome studies in rhesus macaques are currently impaired by the lack of a high-quality genome (Vallender 2011). However, once one becomes available, it should be possible to economically survey colonies of rhesus macaques and identify pairs of animals with mutations in the same Mendelian recessive genes. Directed breeding could then be used to develop nonhuman primate models of human genetic disease. Such animals will be useful for preclinical therapeutic tests.

Acknowledgments

This work was funded by NCRR/ORIP NIH grant R24RR017444.

References

Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J, Humburg P, Iqbal Z, Lunter G, Maller J, Hernandez RD, Melton C, Venkat A, Nobrega MA, Bontrop R, Myers S, Donnelly P, Przeworski M, McVean G. 2012. A fine-scale chimpanzee genetic map from population sequencing. Science 336:193–198.

Baroncelli S, Negri DR, Michelini Z, Cara A. 2008. *Macaca mulatta*, *fascicularis*, and *nemestrina* in AIDS vaccine development. Expert Rev Vaccines 7:1419–1434.

Bosinger SE, Sodora DL, Silvestri G. 2011. Generalized immune activation and innate immune responses in simian immunodeficiency virus infection. Curr Opin HIVAIDS 6:411–418.

Capozzi O, Carbone L, Stanyon RR, Marra A, Yang F, Whelan CW, de Jong PJ, Rocchi M, Archidiacono N. 2012. A comprehensive molecular cytogenetic analysis of chromosome rearrangements in gibbons. Genome Res 22:2520–2528.

Chahroudi A, Bosinger SE, Vanderford TH, Paiardini M, Silvestri G. 2012. Natural SIV hosts: Showing AIDS the door. Science 335:1188–1193.

Chellman GJ, Bussiere JL, Makori N, Martin PL, Ooshima Y, Weinbauer GF. 2009. Developmental and reproductive toxicology studies in nonhuman primates. Birth Defects Res B Dev Reprod Toxicol 86:446–462.

Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.

Daza-Vamenta R, Glusman G, Rowen L, Guthrie B, Geraghty DE. 2004. Genetic divergence of the rhesus macaque major histocompatibility complex. Genome Res 14:1501–15.

Dhand R. 2006. The 'finished' landscape. Nature S1:7.

Duan F, Spindel ER, Li YH, Norgren RB Jr. 2007. Intercenter reliability and validity of the rhesus macaque GeneChip. BMC Genomics 8:61.

Earl D, Bradnam K, St John J, Darling A, Lin D, Fass J, Yu HO, Buffalo V, Zerbino DR, Diekhans M, Nguyen N, Ariyaratne PN, Sung WK, Ning Z, Haimel M, Simpson JT, Fonseca NA, Birol İ, Docking TR, Ho IY, Rokhsar DS, Chikhi R, Lavenier D, Chapuis G, Naquin D, Maillet N, Schatz MC, Kelley DR, Phillippy AM, Koren S, Yang SP, Wu W, Chou WC, Srivastava A, Shaw TI, Ruby JG, Skewes-Cox P, Betegon M, Dimon MT, Solovyev V, Seledtsov I, Kosarev P, Vorobyev D, Ramirez-Gonzalez R, Leggett R, MacLean D, Xia F, Luo R, Li Z, Xie Y, Liu B, Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Yin S, Sharpe T, Hall G, Kersey PJ, Durbin R, Jackman SD, Chapman JA, Huang X, DeRisi JL, Caccamo M, Li Y, Jaffe DB, Green RE, Haussler D, Korf I, Paten B. 2011. Assemblathon 1: A

competitive assessment of de novo short read assembly methods. Genome Res 21:2224–2241.

Ebeling M, Küng E, See A, Broger C, Steiner G, Berrera M, Heckel T, Iniguez L, Albert T, Schmucki R, Biller H, Singer T, Certa U. 2011. Genome-based analysis of the nonhuman primate *Macaca fascicularis* as a model for drug safety assessment. Genome Res 21:1746–1756.

Fox AS, Oler JA, Shelton SE, Nanda SA, Davidson RJ, Roseboom PH, Kalin NH. 2012. Central amygdala nucleus (Ce) gene expression linked to increased trait-like Ce metabolism and anxious temperament in young primates. Proc Natl Acad Sci U S A. 109:18108–18113.

Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, Batzer MA, Bustamante CD, Eichler EE, Hahn MW, Hardison RC, Makova KD, Miller W, Milosavljevic A, Palermo RE, Siepel A, Sikela JM, Attaway T, Bell S, Bernard KE, Buhay CJ, Chandrabose MN, Dao M, Davis C, Delehaunty KD, Ding Y, Dinh HH, Dugan-Rocha S, Fulton LA, Gabisi RA, Garner TT, Godfrey J, Hawes AC, Hernandez J, Hines S, Holder M, Hume J, Jhangiani SN, Joshi V, Khan ZM, Kirkness EF, Cree A, Fowler RG, Lee S, Lewis LR, Li Z, Liu YS, Moore SM, Muzny D, Nazareth LV, Ngo DN, Okwuonu GO, Pai G, Parker D, Paul HA, Pfannkoch C, Pohl CS, Rogers YH, Ruiz SJ, Sabo A, Santibanez J, Schneider BW, Smith SM, Sodergren E, Svatek AF, Utterback TR, Vattathil S, Warren W, White CS, Chinwalla AT, Feng Y, Halpern AL, Hillier LW, Huang X, Minx P, Nelson JO, Pepin KH, Qin X, Sutton GG, Venter E, Walenz BP, Wallis JW, Worley KC, Yang SP, Jones SM, Marra MA, Rocchi M, Schein JE, Baertsch R, Clarke L, Csürös M, Glasscock J, Harris RA, Havlak P, Jackson AR, Jiang H, Liu Y, Messina DN, Shen Y, Song HX, Wylie T, Zhang L, Birney E, Han K, Konkel MK, Lee J, Smit AF, Ullmer B, Wang H, Xing J, Burhans R, Cheng Z, Karro JE, Ma J, Raney B, She X, Cox MJ, Demuth JP, Dumas LJ, Han SG, Hopkins J, Karimpour-Fard A, Kim YH, Pollack JR, Vinar T, Addo-Quaye C, Degenhardt J, Denby A, Hubisz MJ, Indap A, Kosiol C, Lahn BT, Lawson HA, Marklein A, Nielsen R, Vallender EJ, Clark AG, Ferguson B, Hernandez RD, Hirani K, Kehrer-Sawatzki H, Kolb J, Patil S, Pu LL, Ren Y, Smith DG, Wheeler DA, Schenck I, Ball EV, Chen R, Cooper DN, Giardine B, Hsu F, Kent WJ, Lesk A, Nelson DL, O'Brien WE, Prüfer K, Stenson PD, Wallace JC, Ke H, Liu XM, Wang P, Xiang AP, Yang F, Barber GP, Haussler D, Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 108:1513–1518.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech 29:644–652.

Havlak P, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Weinstock GM, Gibbs RA. 2004. The Atlas genome assembly system. Genome Res 14:721–732.

Hewitson L. 2004. Primate models for assisted reproductive technologies. Reproduction 128:293–9.

Higashino A, Sakate R, Kameoka Y, Takahashi I, Hirata M, Tanuma R, Masui T, Yasutomi Y, Osada N. 2012. Whole-genome sequencing and analysis of the Malaysian cynomolgus macaque (*Macaca fascicularis*) genome. Genome Biol 13:R58.

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. Nature 431:931–945.

Karere GM, Froenicke L, Millon L, Womack JE, Lyons LA. 2008. A high-resolution radiation hybrid map of rhesus macaque chromosome 5 identifies rearrangements in the genome assembly. Genomics 92:210–218.

Kehrer-Sawatzki H, Cooper DN. 2008. Molecular mechanisms of chromosomal rearrangement during primate evolution. Chromosome Res 16:41–56

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20:265–272.

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, Wang Z, Chinwalla AT, Minx P, Mitreva M, Cook L, Delehaunty KD, Fronick C, Schmidt H, Fulton LA, Fulton RS, Nelson JO, Magrini V, Pohl C, Graves TA, Markovic C, Cree A, Dinh HH, Hume J, Kovar CL, Fowler GR, Lunter G, Meader S, Heger A, Ponting CP, Marques-Bonet T, Alkan C, Chen L, Cheng Z, Kidd JM, Eichler EE, White S, Searle S, Vilella AJ, Chen Y, Flicek P, Ma J, Raney B, Suh B, Burhans R, Herrero J, Haussler D, Faria R, Fernando O, Darré F, Farré D, Gazave E, Oliva M, Navarro A, Roberto R, Capozzi O, Archidiacono N, Della Valle G, Purgato S, Rocchi M, Konkel MK, Walker JA, Ullmer B, Batzer MA, Smit AF, Hubley R, Casola C, Schrider DR, Hahn MW, Quesada V, Puente XS, Ordoñez GR, López-Otín C, Vinar T, Brejova B, Ratan A, Harris RS, Miller W, Kosiol C, Lawson HA, Taliwal V, Martins AL, Siepel A, Roychoudhury A, Ma X, Degenhardt J, Bustamante CD, Gutenkunst RN, Mailund T, Dutheil JY, Hobolth A, Schierup MH, Ryder OA, Yoshinaga Y, de Jong PJ, Weinstock GM, Rogers J,

Mardis ER, Gibbs RA, Wilson RK. 2011. Comparative and demographic analysis of orang-utan genomes. Nature 469:529–533.

Ma D, Jasinska A, Kristoff J, Grobler JP, Turner T, Jung Y, Schmitt C, Raehtz K, Feyertag F, Martinez Sosa N, Wijewardana V, Burke DS, Robertson DL, Tracy R, Pandrea I, Freimer N, Apetrei C. 2013. SIVagm infection in wild African green monkeys from South Africa: Epidemiology, natural history, and evolutionary considerations. PLoS Pathog 9: e1003011.

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, Albers CA, Zhang ZD, Conrad DF, Lunter G, Zheng H, Ayub Q, DePristo MA, Banks E, Hu M, Handsaker RE, Rosenfeld JA, Fromer M, Jin M, Mu XJ, Khurana E, Ye K, Kay M, Saunders GI, Suner MM, Hunt T, Barnes IH, Amid C, Carvalho-Silva DR, Bignell AH, Snow C, Yngvadottir B, Bumpstead S, Cooper DN, Xue Y, Romero IG; 1000 Genomes Project Consortium, Wang J, Li Y, Gibbs RA, McCarroll SA, Dermitzakis ET, Pritchard JK, Barrett JC, Harrow J, Hurles ME, Gerstein MB, Tyler-Smith C. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. Science 335:823–828.

Magness CL, Fellin PC, Thomas MJ, Korth MJ, Agy MB, Proll SC, Fitzgibbon M, Scherer CA, Miner DG, Katze MG, Iadonato SP. 2005. Analysis of the *Macaca mulatta* transcriptome and the sequence divergence between Macaca and human. Genome Biol 6:R60.

Marques-Bonet T, Ryder OA, Eichler EE. 2009. Sequencing primate genomes: What have we learned? Annu Rev Genomics Hum Genet 10:355–386.

Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC. 2000. A whole-genome assembly of Drosophila. Science 287:2196–2204.

Nagy A, Hegyi H, Farkas K, Tordai H, Kozma E, Bányai L, Patthy L. 2008. Identification and correction of abnormal, incomplete, and mispredicted proteins in public databases. BMC Bioinformatics 9:353.

Nagy A, Szlámal G, Szarkal E, Trexler M, Bányai L, Patthy L. 2011. Reassessing domain architecture evolution of metazoan proteins: Major impact of gene prediction errors. Genes 2:449–501.

Narzisi G, Mishra B. 2011. Comparing de novo genome assembly: The long and short of it. PLoS ONE 6:e19175.

Nie W. 2012. Molecular cytogenetic studies in strepsirrhine primates, Dermoptera and Scandentia. Cytogenet Genome Res 137:246–258.

Norgren RB. 2012. Using NextGen sequencing to identify the molecular basis for genetic disease: An evolutionary perspective. Methods in Next Generation Sequencing 1:1–9.

de Oliveira EH, Neusser M, Müller S. 2012. Chromosome evolution in new world monkeys (Platyrrhini). Cytogenet Genome Res 137:259–272.

Perry GH, Reeves D, Melsted P, Ratan A, Miller W, Michelini K, Louis EE Jr, Pritchard JK, Mason CE, Gilad Y. 2012. A genome sequence resource for the aye-aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar. Genome Biol Evol 4:126–135.

Pipes L, Li S, Bozinoski M, Palermo R, Peng X, Blood P, Kelly S, Weiss JM, Thierry-Mieg J, Thierry-Mieg D, Zumbo P, Chen R, Schroth GP, Mason CE, Katze MG. 2013. The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics. Nucleic Acids Res 41:D906–D914.

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, Cagan A, Theunert C, Casals F, Laayouni H, Munch K, Hobolth A, Halager AE, Malig M, Hernandez-Rodriguez J, Hernando-Herraez I, Prüfer K, Pybus M, Johnstone L, Lachmann M, Alkan C, Twigg D, Petit N, Baker C, Hormozdiari F, Fernandez-Callejo M, Dabad M, Wilson ML, Stevison L, Camprubí C, Carvalho T, Ruiz-Herrera A, Vives L, Mele M, Abello T, Kondova I, Bontrop RE, Pusey A, Lankester F, Kiyang JA, Bergl RA, Lonsdorf E, Myers S, Ventura M, Gagneux P, Comas D, Siegismund H, Blanc J, Agueda-Calpena L, Gut M, Fulton L, Tishkoff SA, Mullikin JC, Wilson RK, Gut IG, Gonder MK, Ryder OA, Hahn BH, Navarro A, Akey JM, Bertranpetit J, Reich D, Mailund T, Schierup MH, Hvilsom C, Andrés AM, Wall JD, Bustamante CD, Hammer MF, Eichler EE, Marques-Bonet T. 2013. Great ape genetic diversity and population history. Nature (doi:10.1038/nature12228). Available online (http://www.nature.com/nature/journal/vaop/ncurrent/full/nature12228.html), accessed on July 3, 2013.

Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, Knight JR, Mullikin JC, Meader SJ, Ponting CP, Lunter G, Higashino S, Hobolth A, Dutheil J, Karakoç E, Alkan C, Sajjadian S, Catacchio CR, Ventura M, Marques-Bonet T, Eichler EE, André C, Atencia R, Mugisha L, Junhold J, Patterson N, Siebauer M, Good JM, Fischer A, Ptak SE, Lachmann M, Symer DE, Mailund T, Schierup MH, Andrés AM, Kelso J, Pääbo S. 2012. The bonobo genome compared with the chimpanzee and human genomes. Nature 486:527–531.

Roberto R, Misceo D, D'Addabbo P, Archidiacono N, Rocchi M. 2008. Refinement of macaque synteny arrangement with respect to the official rheMac2 macaque sequence assembly. Chromosome Res 16:977–985.

Rogers J, Garcia R, Shelledy W, Kaplan J, Arya A, Johnson Z, Bergstrom M, Novakowski L, Nair P, Vinson A, Newman D, Heckman G, Cameron J. 2006. An initial genetic linkage map of the rhesus macaque (*Macaca mulatta*) genome using human microsatellite loci. Genomics 87:30–38.

Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, Treangen TJ, Schatz MC, Delcher AL, Roberts M, Marçais G, Pop M, Yorke JA. 2012. GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Res 22:557–567.

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, McCarthy S, Montgomery SH, Schwalie PC, Tang YA, Ward MC, Xue Y, Yngvadottir B, Alkan C, Andersen LN, Ayub Q, Ball EV, Beal K, Bradley BJ, Chen Y, Clee CM, Fitzgerald S, Graves TA, Gu Y, Heath P, Heger A, Karakoc E, Kolb-Kokocinski A, Laird GK, Lunter G, Meader S, Mort M, Mullikin JC, Munch K, O'Connor TD, Phillips AD, Prado-Martinez J, Rogers AS, Sajjadian S, Schmidt D, Shaw K, Simpson JT, Stenson PD, Turner DJ, Vigilant L, Vilella AJ, Whitener W, Zhu B, Cooper DN, de Jong P, Dermitzakis ET, Eichler EE, Flicek P, Goldman N, Mundy NI, Ning Z, Odom DT, Ponting CP, Quail MA, Ryder OA, Searle SM, Warren WC, Wilson RK, Schierup MH, Rogers J, Tyler-Smith C, Durbin R. 2012. Insights into hominid evolution from the gorilla genome sequence. Nature 483:169–175.

Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092.

Shackman AJ, Fox AS, Oler JA, Shelton SE, Davidson RJ, Kalin NH. 2013. Neural mechanisms underlying heterogeneity in the presentation of anxious temperament. Proc Natl Acad Sci U S A 110:6145–6150.

Shedlock DJ, Silvestri G, Weiner DB. 2009. Monkeying around with HIV vaccines: Using rhesus macaques to define 'gatekeepers' for clinical trials. Nat Rev Immunol 9:717–728.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Res 19:1117–1123.

Sparman ML, Ramsey CM, Thomas CM, Mitalipov SM, Fanton JW, Maginnis GM, Stouffer RL, Wolf DP. 2007. Evaluation of the vervet (*Clorocebus aethiops*) as a model for the assisted reproductive technologies. Am J Primatol 69:917–929.

Spindel ER, Pauley MA, Jia Y, Gravett C, Thompson SL, Boyle NF, Ojeda SR, Norgren RB Jr. 2005. Leveraging human genomic information to identify nonhuman primate sequences for expression array development. BMC Genomics 6:160.

Stanyon R, Rocchi M, Bigoni F, Archidiacono N. 2012. Evolutionary molecular cytogenetics of catarrhine primates: Past, present and future. Cytogenet Genome Res 137:273–284.

Torkamani A, Pham P, Libiger O, Bansal V, Zhang G, Scott-Van Zeeland AA, Tewhey R, Topol EJ, Schork NJ. 2012. Clinical implications of human population differences in genome-wide rates of functional genotypes. Front Genet 3:211.

Vallender EJ. 2009. Bioinformatic approaches to identifying orthologs and assessing evolutionary relationships. Methods 49:50–55.

Vallender EJ. 2011. Expanding whole exome resequencing into non-human primates. Genome Biol 12:R87.

Vinson A, Curran JE, Johnson MP, Dyer TD, Moses EK, Blangero J, Cox LA, Rogers J, Havill LM, Vandeberg JL, Mahaney MC. 2011. Genetical genomics of Th1 and Th2 immune response in a baboon model of atherosclerosis risk factors. Atherosclerosis 217:387–394.

Watson KK, Platt ML. 2012. Of mice and monkeys: Using non-human primate models to bridge mouse- and human-based investigations of autism spectrum disorders. J Neurodev Disord 4:21.

Wheelan SJ, Church DM, Ostell JM. 2001. Spidey: A tool for mRNA-to-genomic alignments. Genome Res 11:1952–1957.

Wu TD, Watanabe CK. 2005. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21:1859–1875.

Yan G, Zhang G, Fang X, Zhang Y, Li C, Ling F, Cooper DN, Li Q, Li Y, van Gool AJ, Du H, Chen J, Chen R, Zhang P, Huang Z, Thompson JR, Meng Y, Bai Y, Wang J, Zhuo M, Wang T, Huang Y, Wei L, Li J, Wang Z, Hu H, Yang P, Le L, Stenson PD, Li B, Liu X, Ball EV, An N, Huang Q, Zhang Y, Fan W, Zhang X, Li Y, Wang W, Katze MG, Su B, Nielsen R, Yang H, Wang J, Wang X, Wang J. 2011. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. Nature Biotechnology 29:1019–1023.

Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829.

Zhang X, Goodsell J, Norgren RB. 2012 Limitations of the rhesus macaque draft genome assembly and annotation. BMC Genomics 13:206.

Zhou L, Pertea M, Delcher AL, Florea L. 2009. Sim4cc: A cross-species spliced alignment program. Nucleic Acids Res 37:e80.