MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 703–710 doi: 10.1111/j.1755-0998.2011.02999.x

Gene discovery in the Antarctic fur seal (*Arctocephalus gazella*) skin transcriptome

JOSEPH I. HOFFMAN
Department of Animal Behaviour, University of Bielefeld, Postfach 100131, 33501 Bielefeld, Germany

Abstract

Next-generation sequencing provides a powerful new approach for developing functional genomic tools for nonmodel species, helping to narrow the gap between studies of model organisms and those of natural populations. Consequently, massively parallel 454 sequencing was used to characterize a normalized cDNA library derived from skin biopsy samples of twelve Antarctic fur seal (*Arctocephalus gazella*) individuals. Over 412 Mb of sequence data were generated, comprising 1.4 million reads of average length 286 bp. *De novo* assembly using Newbler 2.3 yielded 156 contigs plus 22 869 isotigs, which in turn clustered into 18 576 isogroups. Almost half of the assembled transcript sequences showed significant similarity to the nr database, revealing a functionally diverse array of genes. Moreover, 97.9% of these mapped to the dog (*Canis lupis familiaris*) genome, with a strong positive relationship between the number of sequences locating to a given chromosome and the length of that chromosome in the dog indicating a broad genomic distribution. Average depth of coverage was also almost 20-fold, sufficient to detect several thousand putative microsatellite loci and single nucleotide polymorphisms. This study constitutes an important step towards developing genomic resources with which to address consequential questions in pinniped ecology and evolution. It also supports an earlier but smaller study showing that skin tissue can be a rich source of expressed genes, with important implications for studying the genomics not only of marine mammals, but also more generally of species that cannot be destructively sampled.

Keywords: dog (*Canis lupis familiaris*), expressed sequence tags, marine mammal, microsatellite, pinniped, single nucleotide polymorphism, 454 sequencing

Received 16 November 2010; revision received 7 January 2011; accepted 17 January 2011

Introduction

Next-generation sequencing technologies are rapidly transforming the study of evolutionary and population genetics, offering an unprecedented opportunity to characterize genetic variation both in model and nonmodel organisms (Rokas & Abbot 2009). Roche's 454 platform (Margulies *et al.* 2005) has been particularly instrumental in respect of the latter because it generates reads of a sufficient length, quality and number to allow accurate *de novo* sequence assembly, a prerequisite for transcriptome characterization. However, despite 454 sequencing being cheaper and far less labour intensive than the traditional approach of cloning and Sanger sequencing expressed sequence tags (ESTs), it has not yet been widely applied to nonmodel species. As a result, genomic resources are still largely lacking for many ecologically compelling natural systems.

Marine mammals are an interesting case in point, comprising a large and charismatic animal assemblage, but one for which relatively few genomic resources are currently available. Thus, although complete mitochondrial genomes have been published for a number of marine mammals (Arnason *et al.* 2006) and several species are also listed in the 'Genome 10K proposal' (Genome 10K Community 2009), a full-genome sequence has only so far been compiled for the bottlenose dolphin *Tursiops truncatus* (http://www.hgsc.bcm.tmc.edu/project-species-m-Dolphin.hgsc?pageLocation=Dolphin). ESTs have also been cloned and Sanger sequenced from this species (Mancia *et al.* 2007) as well as from the North Atlantic right whale *Eubalaena glacialis* (Ierardi *et al.* 2009), but to my knowledge no 454-based marine mammal transcriptome studies have been published to date.

Even fewer such resources appear to have been developed for pinnipeds (meaning 'wing footed'), a group of marine mammals that includes seals, sea lions and walruses. Pinnipeds are of great interest from an evolutionary standpoint, partly because they make good models for studying adaptation to a semi-aquatic lifestyle, but also because they are among the most sexually dimorphic and polygynous of all mammals (Boness 1991). However, searches of the National Centre for Biotechnology Information (NCBI) nucleotide sequence database conducted

Correspondence: Joseph I. Hoffman, Fax: +49 521 1062998;
E-mail: joseph.hoffman@uni-bielefeld.de

separately for each of the 33 extant pinniped species on the 6 November 2010 recovered a total of only 5068 pinniped-specific sequences, of which 75.0% were either mitochondrial control region/cytochrome B or microsatellite-containing clones. Parallel searches of NCBI's EST database did not reveal any pinniped ESTs.

The Antarctic fur seal (*Arctocephalus gazella*) is emerging as an important model pinniped system. This species is highly sexually dimorphic and breeds in crowded rookeries where it exhibits resource defence polygyny (McCann 1980). On Bird Island, South Georgia, a colony of Antarctic fur seals has been studied since the 1950s with an aerial walkway providing unprecedented access for tissue sampling and the collection of detailed daily behavioural observations (Doidge *et al.* 1984). Genetic analysis using nine hypervariable microsatellite loci has shown that most if not all pups are conceived on land (Hoffman *et al.* 2003) and hence that lifetime reproductive profiles can be constructed for virtually every individual sighted ashore. Heterozygosity at the same panel of markers has also been shown to correlate with virtually every aspect of male reproductive success measured in this species including the ability to hold a territory both within and across seasons (Hoffman *et al.* 2004), body size (Hoffman *et al.* 2010a) and attractiveness to females (Hoffman *et al.* 2007a). However, because only a handful of anonymous microsatellites have been analysed so far, the mechanisms underlying these associations remain unclear. Consequently, it would be advantageous to develop additional markers, especially if these could be targeted within functionally annotated genes.

This article reports the 454 sequencing of a cDNA library derived from the skin samples of twelve Antarctic fur seal individuals. Skin was targeted as a source of mRNA both because it is relatively straightforward to biopsy sample fur seals (Gemmell & Majluf 1997) and because a previous study was able to isolate functionally diverse sequences from the skin of North Atlantic right whales (Ierardi *et al.* 2009). The aims of this study were to generate an annotated sequence database for the Antarctic fur seal, thereby also providing a preliminary genomic resource for pinnipeds in general, and to identify putative molecular markers with which to further explore links between genotype and fitness in this and closely related species.

Materials and methods

Tissue sampling

Skin biopsy samples were collected from six adult males, two adult females and four pups at Freshwater Beach on Bird Island, South Georgia (54°00'S, 38°02'W) during the austral summer of 2008/2009. Adult males were sampled using a remote biopsy dart system (Gemmell & Majluf 1997) and piglet ear-notching pliers were used to collect a small skin sample from the interdigital margin of the foreflipper of adult females and pups. All sampling equipment was sterilized using 95% ethanol between uses. Skin samples were immediately transferred to RNAlater® and stored individually at −20 °C for up to 1 day before being placed in a −80 °C freezer for transport back to the UK.

RNA isolation

Approximately 10 mg of each tissue sample was disrupted and homogenized by bead milling within a TissueLyserII (Qiagen). Total RNA was then extracted using a Qiagen RNeasy® mini kit following the manufacturer's recommended protocols, with an optional on-column DNAse digestion step included. The resulting RNA pellets were individually resuspended in 50 μl of RNAse-free water (Ambion) and quantified using a NanoDrop™ 2000 spectrophotometer (Thermo Scientific). RNA quantity and quality were also assessed visually by running a fraction of each isolate on a 1% agarose gel.

cDNA library construction and 454 sequencing

RNA extracts were pooled in equimolar amounts and complimentary DNA synthesized using a SMART cDNA kit (Clontech) following the manufacturer's recommended protocol. To decrease the abundance of highly abundant transcripts and facilitate the discovery of genes expressed at low levels, normalization was performed using a TRIMMER cDNA normalization kit (Evrogen). The resulting cDNA fraction was then PCR-amplified using PCR Advantage® two polymerase (Clontech) with the following thermal profile: initial denaturation for 1 min at 95 °C followed by 19 cycles of 95 °C for 7 s, 65 °C for 20 s and 72 °C for 3 min. After purification using QIAquick PCR Purification Kit (Qiagen), the normalized cDNA was then sheared using a Covaris™ S2 (Applied Biosystems) and 0.5 μg sequenced using established protocols (Margulies *et al.* 2005) on a Roche GS-FLX DNA sequencer (Roche Diagnostic) at the Edinburgh GenePool (Edinburgh University, UK). The resulting sequence data were archived in GenBank (SRA: ERP000497).

Sequence assembly

Individual reads, masked for SMART oligonucleotide adaptors and trimmed of poly-A tails, were assembled using Roche Newbler assembler version 2.3. This program has a cDNA option designed for *de novo* transcriptome assembly (see Genome sequencer FLX sys-

tem software manual part C, 2009, for a detailed description). Briefly, Newbler works by first assembling overlapping reads into contigs, which can be broadly thought of as exons, although untranslated regions (5′ UTR and 3′ UTR) and introns (in the case of primary transcripts) may also be present among the reads. Contigs containing reads that imply connections among them are next clustered into isogroups, each representing a collection of transcripts of a given gene (see Fig. 1 for a hypothetical example). The assembler finally traverses the various paths among the contigs within an isogroup to produce a set of isotigs. Different isotigs from a given isogroup can be inferred as splice variants.

BLAST searches and functional annotation

To determine homology to known genes, basic local alignment search tool (BLAST) searches (Altschul *et al.* 1990) with an e-value cut-off of $1e^{-4}$ were used to query contig and isotig sequences against the GenBank nonredundant (nr) database. The program Annot8r was then used to conduct a further set of BLAST searches against a subset of sequences with known gene ontology (GO) annotation. This analysis was carried out using a default bit score cutoff of 55 and allowing up to 50 matches per isotig. However, to provide a simplified overview of the ontology content, only hits against the GO-Slim subset are presented.

Exploring sequence homology with the dog genome

Using BLAST searches, Hoffman *et al.* (2010b) recently demonstrated significant homology between the clone sequences of 94% of 76 anonymous pinniped microsatellite loci and unique locations within the dog (*Canis lupus familiaris*) genome. Consequently, two approaches were employed to further explore sequence homology between the fur seal and dog. First, individual 454 reads were mapped to the dog genome (build 2.0) using Roche gsMapper version 2.3. The dog genome sequences in FASTA format were obtained from ftp://ftp.ncbi.nih.gov/genomes/Canis_familiaris and comprised chromosomes 1–38 and X, but unfortunately no sequence data were available for the Y chromosome. Second, a final set of BLAST searches were conducted, this time using the assembled fur seal contigs/isotigs as query sequences, an e-value cut-off of $1e^{-4}$ and restricting the search set to canine sequences.

Molecular marker discovery

Finally, searches were carried out for microsatellite motifs using the program SSRIT (Temnykh *et al.* 2001) to identify sequences containing perfect di-, tri- and tetranucleotide repeats with a minimum length of five repeat units. SNP detection was conducted using the SWAP454 pipeline (Brockman *et al.* 2008). Because sequence quality scores tend to be lower towards the ends of reads and in the vicinity of long homopolymers, SWAP454 incorporates a novel *phred*-based quality score into the SNP-calling algorithm that has been shown to be effective in reducing the occurrence of false positives. The program starts by mapping the raw reads back to the assembled contigs/isotigs and then determines, while taking into account an error model for the 454 data, which positions are called as SNPs according to two user-specified thresholds. The first of these thresholds, 'MIN_RATIO' corresponds to the percentage of reads that differ from the reference sequence at a given position and the second, 'MIN_READS' to the number of copies present of the minor allele. For this analysis, two criteria were applied: a 'strict' criterion to minimize the possibility of false positives arising from sequencing error and a 'relaxed' criterion to maximize the discovery of relatively infrequent alleles. For the former, MIN_RATIO was set to 0.33 and MIN_READS to 8, which is more stringent than the threshold applied by Vera *et al.* (2008), and for the latter MIN_RATIO was set to 0.1 and MIN_READS to 2. A subset of the resulting putative SNPs was verified by visual inspection within the program Tablet v1 (Milne *et al.* 2010).

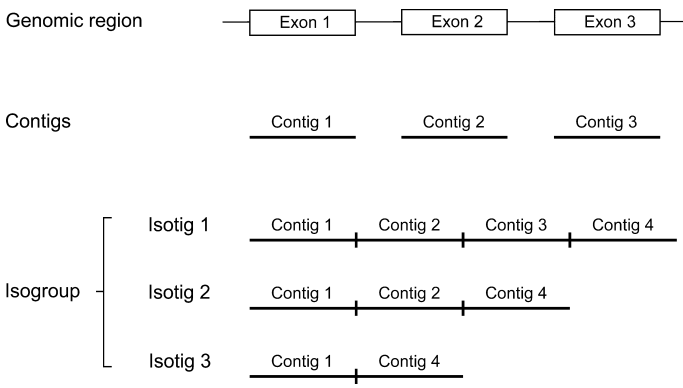

Fig. 1 Schematic of a hypothetical Newbler assembly, illustrating relationships among exons, contigs, isotigs and isogroups (see the Materials and methods section for definitions of these terms).

Table 1 Summary statistics for the Antarctic fur seal 454 sequence assembly

Reads	Number	1 443 397
	Mean length	286
	Total base pairs (bp)	412 933 200
Singletons	Number	170 075
	Mean length	187
Contigs	Number	156
	Mean length	1069
	Mean coverage	37.8
Isotigs	Number	22 869
	Mean length	854
	Mean coverage	19.4

Results and discussion

Sequencing and assembly

A normalized fur seal cDNA library was subject to a full 454 sequencing run, yielding a total of 1 443 397 reads with a mean length of 286 nucleotides and totalling over 400 Mb (Table 1, see Fig. 2a for the length distribution). These were assembled *de novo* into 156 contigs plus 22 869 isotigs, which for simplicity are hereafter collectively referred to as isotigs (total n = 23 025). Mean isotig length was 854 bp, and the average depth of coverage was almost 20-fold (Table 1, Fig. 2b, c). Newbler further clustered these sequences into 18 576 isogroups. A total of 170 075 singleton reads remained unassembled and were not included in subsequent analyses. This is because, although singletons probably contain useful information about transcripts that are expressed at very low levels, they can also be prone to artefacts arising during cDNA synthesis and sequencing or from contamination (Mayer *et al*. 2009).

BLAST searches and functional annotation

Sequence similarity searches to the nr database using BLAST produced matches for 10 825 isotig sequences (47.0%). Although an e-value threshold of $1e^{-4}$ was used for this analysis, the majority of matches yielded far lower e-values (Fig. 3a). As observed by Wang *et al*. (2009), the proportion of matching isotigs correlated significantly with sequence length, increasing from 31.2% for isotigs shorter than 500 bp (n = 7390) to 85.4% for isotigs longer than 2000 bp (n = 1276). Reassuringly, the majority of sequences (96.9%) showed top matches with other vertebrate species (Fig. 3b). The remaining matches involved mostly pathogenic bacteria (2.2%), consistent with a previous study that isolated a diverse bacterial assemblage from fight wounds and lesions in adult males of this species (Baker & McCann 1989), and were removed from subsequent analyses. Among the sequences showing significant similarity to vertebrate species, 99.1% of the top matches were to mammals (Fig. 3c) and these most frequently comprised the dog followed by the panda (Fig. 3d).

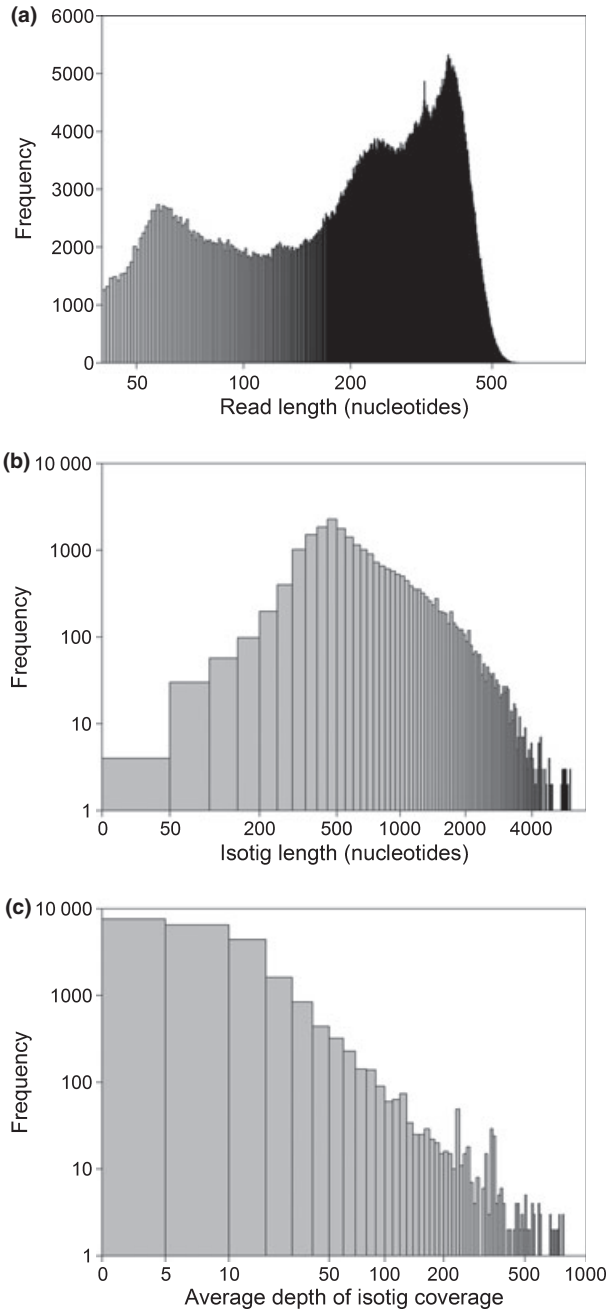

Fig. 2 Frequency distributions of (a) read lengths, (b) isotig lengths and (c) average isotig depth of coverage. Variable bar widths are due to the use of logarithmic horizontal axes. Note that logarithmic vertical axes have also been used in panels (b) and (c).

Using standard GO annotation, the isotigs were next classified according to three major functional categories: biological process, molecular function and cellular com-

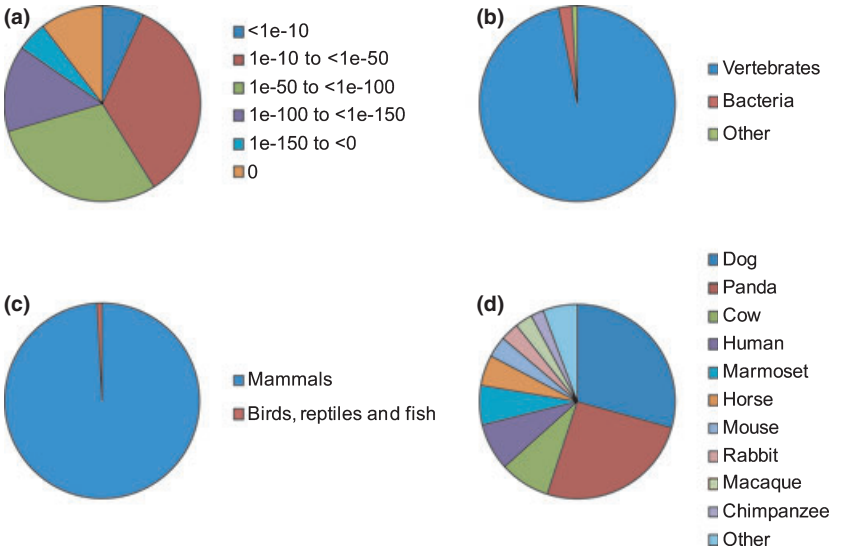

Fig. 3 Characteristics of top basic local alignment search tool hits (n = 10 825 isotigs). (a) E-value distribution; (b–d) taxonomic distribution summarized respectively for all species, vertebrates only and mammals only.

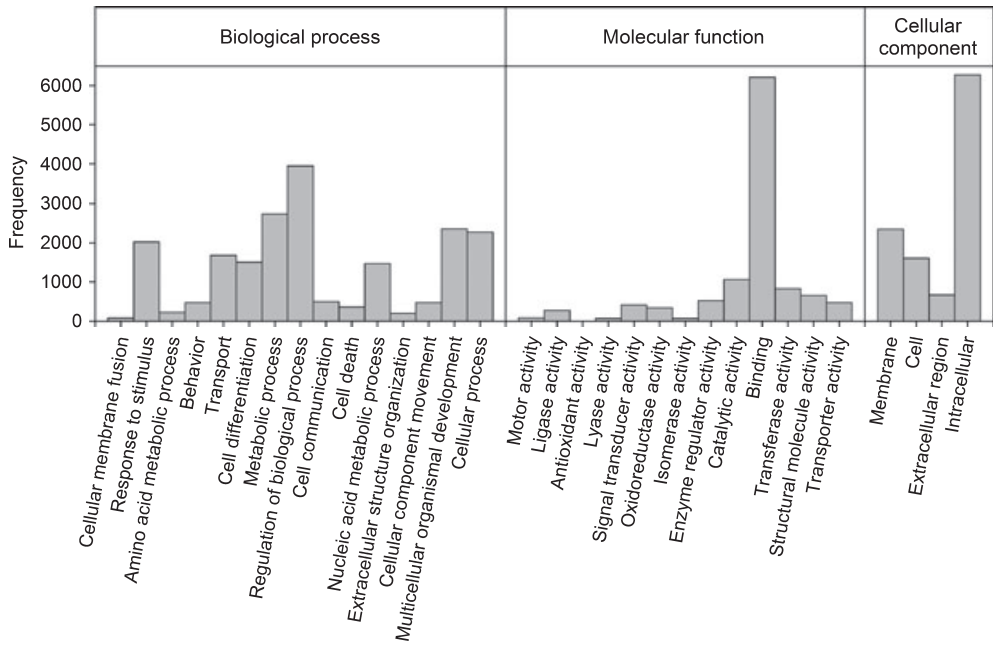

Fig. 4 Gene ontology classification (n = 7668 isotigs), summarized for three major categories: biological process, molecular function and cellular component.

ponent. Biological processes made up the majority of GO-Slim designations (n = 20 311) followed by molecular function (n = 11 035) and then cellular component (n = 10 903). Figure 4 shows that the annotated sequences are functionally diverse, supporting a previous but smaller study of North Atlantic right whales (Ierardi *et al.* 2009) which concluded that skin can be a rich source of expressed genes. This finding could be important for the future development of genomic applications in marine mammals given that it is difficult even to obtain skin samples from the majority of species. It could also have implications for investigators wishing to apply similar techniques in the context of conservation biology, where again constraints are often present on the types of tissues that can be collected for genetic analysis. One caveat to the above is that a small number of categories were found to be relatively over-represented, including 'regulation of cellular biological processes', 'metabolism' and 'binding'. However, this is also consistent with several previous studies, including ones that

have utilized multiple types of tissue (e.g. Ierardi *et al.* 2009; Schwarz *et al.* 2009; Wang *et al.* 2010).

Comparative genomics with the dog

Seals and dogs diverged approximately 40 million years ago, a sufficiently short timescale to allow Slade *et al.* (1994) to PCR-amplify and align conserved exonic regions from several genes in both the dog and a number of pinniped species. More recently, significant homology to the dog was also observed for the full-length clone sequences of 50% of 76 anonymous pinniped microsatellites, with this proportion increasing to 94% when searches were carried out using the longest, least repetitive sequence from one side of the marker (Hoffman *et al.* 2010b). These loci were found to be distributed across 33 different chromosomes in the dog, with a maximum of five locating to any one chromosome and four putatively X-linked markers all mapping to the X chromosome. To shed light on the genomic distribution of the 454 sequence data generated in this study, both unassembled reads and isotig sequences were mapped to the dog genome. A total of 679 558 individual reads (47.1%) successfully mapped, once again revealing a fairly even genomic distribution evidenced by a strong positive relationship between the number of reads mapping to a given chromosome and the length of that chromosome in the dog (Fig. 5a, $r^2 = 0.70$, $n = 39$, $P < 0.0001$). Furthermore, almost all of the isotigs successfully mapped to the dog genome ($n = 22\,541$, 97.9%) yielding an even stronger equivalent relationship (Fig. 5b, $r^2 = 0.78$, $n = 39$, $P < 0.0001$). This bodes well for future studies because it allows inferences to be drawn about the genomic distribution of the sequences and, by implication, any genetic markers residing within them. Moreover, although full-genome sequencing of the fur seal is not envisaged in the immediate future, homology with the dog could potentially be exploited to facilitate the assembly process.

Molecular marker discovery

Searches for perfect microsatellites with at least five repeat units identified 2271 loci, 1871 (82.4%) of which comprised dinucleotides, 301 (13.3%) trinucleotides and 99 (4.4%) tetranucleotides (available via Dryad, doi: 10.5061/dryad.8268). These were located within 1939 different isotigs, of which 864 (44.6%) were functionally annotated with respect to the nr database and 1834 (94.6%) mapped to known regions in the dog genome. Using a 'strict' criterion in which SNPs were accepted when at least 33% of the reads differed from the reference sequence and the minor allele was observed at least eight times, a total of 642 SNPs were identified. These were distributed across 470 isotigs, of which 285 (60.6%) were annotated and 448 (95.3%) mapped to the dog. As anticipated, the number of SNPs that were accepted by the program increased by almost an order of magnitude (to $n = 6261$) when a 'relaxed' criterion was applied that required only 10% of the reads to differ from the reference sequence and two copies of the minor allele to be present.

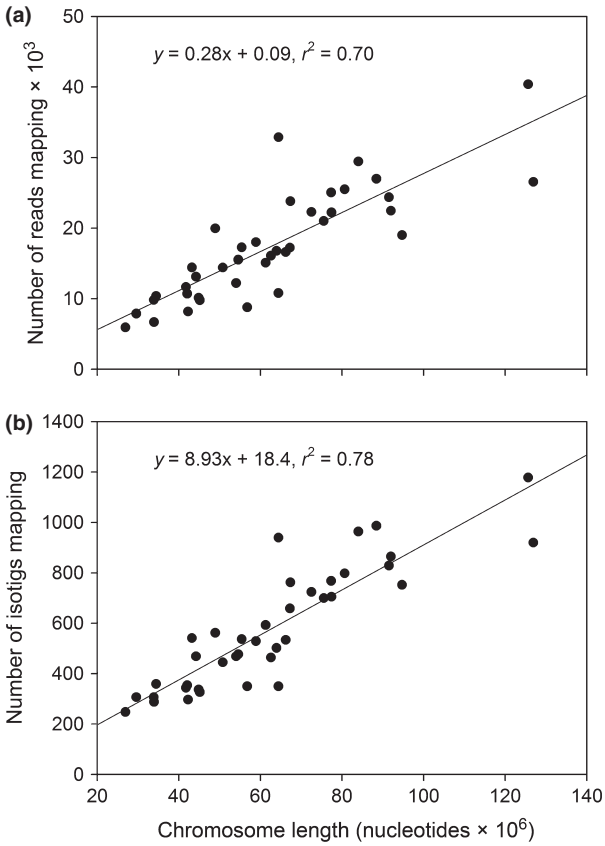

Fig. 5 Relationships between the numbers of (a) 454 reads and (b) assembled isotigs mapping to each chromosome within the dog (*Canis lupus familiaris*) genome and chromosome length (see Materials and methods for details). The Y chromosome is not included because no sequence data were available.

The genetic markers identified above should provide a valuable resource for future genomic and population genetic studies of fur seals and related pinniped species. Both types of markers have distinct advantages and disadvantages (Morin *et al.* 2004). On the one hand, microsatellites tend to be more informative than SNPs, making them well suited to detecting further associations between genotype and fitness (Amos *et al.* online early). The proven ability of microsatellites isolated from fur seals and sea lions to cross-amplify across the Pinnipedia (Hoffman 2007; Hoffman *et al.* 2007b) could also potentially be exploited to allow comparative analyses. However, microsatellite genotyping is labour intensive and careful scrutiny is required to minimize the impact of

human error (Hoffman & Amos 2005), ultimately limiting the number of markers that can be realistically deployed.

SNPs on the other hand are individually less informative than microsatellites and are more species-specific in nature. However, they are far better suited to fully automated medium- to high-throughput genotyping strategies that in principle allow hundreds or even thousands of markers to be genotyped on a large scale and with minimal error. For example, Applied Biosystem's SNPlex and Beckman's SNPStream systems allow up to 48 SNPs to be multiplexed, while Illumina's Goldengate assay appears to work well for larger numbers of loci (<1536, Slate *et al*. 2009). Ultra-high-throughput methods capable of screening tens or even hundreds of thousands of SNPs are also envisaged to become increasingly available. Already, both Affymetrix and Illumina offer the opportunity to develop custom SNP chips capable of screening >10 000 SNPs per individual.

Conclusions

In respect of gene discovery, massively parallel 454 sequencing offers numerous advantages over the classical approach of individually cloning and Sanger sequencing EST libraries. Despite this, however, 454 sequencing has not yet been widely applied to natural populations of nonmodel species such as marine mammals. This study sequenced the skin transcriptome of the Antarctic fur seal, characterizing a variety of expressed genes and at the same time strengthening our current understanding of sequence homology between seals and the dog. Significant numbers of putative markers were also identified, providing candidates for future testing and development. Finally, with only around 5000 pinniped sequences deposited to date at NCBI, most of which correspond to just a handful of genes, the data generated by this study should greatly enlarge the existing pinniped sequence database.

Acknowledgements

I am grateful to Jaume Forcada for collecting tissue samples, A. Montazam and D. Cleven for conducting the 454 sequencing and S. Bridgett who provided invaluable assistance with the bioinformatic analyses. I also thank J. Slate for helpful discussions on next-generation sequence assembly and SNP discovery and two anonymous referees for comments that improved the manuscript. This work was funded by a Natural Environment Research Council (NERC) BAS Strategic Alliance Fellowship awarded to JIH. Fieldwork was approved by BAS, and samples were collected and retained under permits issued by the Department for Environment, Food and Rural Affairs and in accordance with the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). The 454 sequencing was undertaken at the GenePool facility, University of Edinburgh.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic logical alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Amos W, Driscoll E, Hoffman JI (online early) Candidate genes versus genomewide associations: which are better for detecting genetic susceptibility to infectious disease? *Proceedings of the Royal Society of London Series B Biological Sciences*, doi: 10.1098/rspb.2010.1920.

Arnason U, Gullberg A, Janke A *et al*. (2006) Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Molecular Phylogenetics and Evolution*, **41**, 345–354.

Baker JR, McCann TS (1989) Pathology and bacteriology of adult male Antarctic fur seals, *Arctocephalus gazella*, dying at Bird Island, South Georgia. *British Veterinary Journal*, **145**, 263–276.

Boness DJ (1991) Determinants of mating systems in the Otariidae (Pinnipedia). In: *The Behaviour of Pinnipeds* (ed. Renouf D), pp. 1–64, Chapman and Hall, New York.

Brockman W, Alvarez P, Young S *et al*. (2008) Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Research*, **18**, 763–770.

Doidge DW, Croxall JP, Baker JR (1984) Density-dependent pup mortality in the Antarctic fur seal *Arctocephalus gazella* at South Georgia. *Journal of Zoology*, **202**, 449–460.

Gemmell NJ, Majluf P (1997) Projectile biopsy sampling of fur seals. *Marine Mammal Science*, **13**, 512–516.

Genome 10K Community of Scientists (2009) A proposal to obtain whole-genome sequence for 10 000 vertebrate species. *Journal of Heredity*, **100**, 659–674.

Genome sequencer FLX system software manual, Software v. 2.3, Part C (2009) GS *De Novo* assembler, GS reference mapper, SFF tools. Roche.

Hoffman JI (2007) A panel of new microsatellite loci for genetic studies of Antarctic fur seals and other otariid species. *Conservation Genetics*, **10**, 989–992.

Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, **14**, 599–612.

Hoffman JI, Boyd IL, Amos W (2003) Male reproductive strategy and the importance of maternal status in the Antarctic fur seal *Arctocephalus gazella*. *Evolution*, **57**, 1917–1930.

Hoffman JI, Boyd IL, Amos W (2004) Exploring the relationship between parental relatedness and male reproductive success in the Antarctic fur seal *Arctocephalus gazella*. *Evolution*, **58**, 2087–2099.

Hoffman JI, Forcada J, Trathan PN, Amos W (2007a) Female fur seals show active choice for males that are heterozygous and unrelated. *Nature*, **445**, 912–914.

Hoffman JI, Wolf JBW, Steinfartz S (2007b) Ten novel dinucleotide microsatellite loci cloned from the Galapagos sea lion (*Zalophus californianus wollebaeki*) are polymorphic in other pinniped species. *Molecular Ecology Notes*, **7**, 103–105.

Hoffman JI, Hanson N, Forcada J, Trathan PN, Amos W (2010a) Getting long in the tooth: a strong positive correlation between canine size and heterozygosity in the Antarctic fur seal *Arctocephalus gazella*. *Journal of Heredity*, **101**, 527–538.

Hoffman JI, Forcada J, Amos W (2010b) Exploring the mechanisms underlying a heterozygosity–fitness correlation for canine size in the Antarctic fur seal *Arctocephalus gazella*. *Journal of Heredity*, **101**, 539–552.

Ierardi JL, Mancia A, McMillan J *et al*. (2009) Sampling the skin transcriptome of the North Atlantic right whale. *Comparative Biochemistry and Physiology Part D*, **4**, 154–158.

Mancia A, Lundqvist ML, Romano TA *et al*. (2007) A dolphin peripheral blood leukocyte cDNA microarray for studies of immune function and stress reactions. *Developmental and Comparative Immunology*, **31**, 520–509.

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

Mayer E, Aglyamova GV, Wang S *et al*. (2009) Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics*, **10**, 219.

McCann TS (1980) Territoriality and breeding behaviour of adult male Antarctic fur seal, *Arctocephalus gazella*. *Journal of Zoology*, **192**, 295–310.

Milne I, Bayer M, Cardle L *et al*. (2010) Tablet -next generation sequence assembly visualization. *Bioinformatics*, **26**, 401–402.

Morin PA, Luikart G, Wayne RK; the SNP workshop group (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, **19**, 208–216.

Rokas A, Abbot P (2009) Harnessing genomics for evolutionary insights. Harnessing genomics for evolutionary insights. *Trends in Ecology and Evolution*, **24**, 192–200.

Schwarz D, Robertson HM, Feder JL *et al*. (2009) Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. *BMC Genomics*, **10**, 633.

Slade RW, Moritz C, Heideman A (1994) Multiple nuclear-gene phylogenies: application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. *Molecular Biology and Evolution*, **11**, 341–356.

Slate J, Gratten J, Beraldi D, Stapley J, Hale M, Pemberton JM (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica*, **136**, 97–107.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Research*, **11**, 1441–1452.

Vera JC, Wheat CW, Fescemyer HW *et al*. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 sequencing. *Molecular Ecology*, **17**, 1636–1647.

Wang W, Wang Y, Zhang Q, Qi Y, Guo D (2009) Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics*, **10**, 465.

Wang X-W, Luan J-B, Li J-M, Bao Y-Y, Zhang C-X, Lui S-S (2010) De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, **11**, 400.

Joe Hoffman is interested in population genetics, animal behaviour and conservation biology. His ongoing collaborations with colleagues from the British Antarctic Survey range from a long-term genetic study of an Antarctic fur seal colony at South Georgia to the analysis of population structure and adaptation in marine invertebrates from the Antarctic Peninsula.

Data accessibility

DNA sequences: Genbank Sequence Read Archive number: ERP000497.

DNA sequence assembly together with a list of isotigs containing microsatellite motifs: Dryad entry doi: 10.5061/dryad.8268.