Review

Cell PRESS

Adaptation Genomics: the next generation

Jessica Stapley[1,*], Julia Reger[1,*], Philine G.D. Feulner[2], Carole Smadja[3], Juan Galindo[1], Robert Ekblom[4], Clair Bennison[1], Alexander D. Ball[1], Andrew P. Beckerman[1] and Jon Slate[1]

[1] Department of Animal & Plant Sciences, University of Sheffield, S10 2TN, UK
[2] Evolutionary Bioinformatics Group, Institute for Evolution and Biodiversity, Westfälische Wilhelms University, Munster, Germany
[3] CNRS- Institute of Evolutionary Sciences (ISEM), University of Montpellier, France
[4] Population Biology and Conservation Biology, Department of Ecology and Evolution, Evolutionary Biology Centre, Uppsala University, Sweden

Understanding the genetics of how organisms adapt to changing environments is a fundamental topic in modern evolutionary ecology. The field is currently progressing rapidly because of advances in genomics technologies, especially DNA sequencing. The aim of this review is to first briefly summarise how next generation sequencing (NGS) has transformed our ability to identify the genes underpinning adaptation. We then demonstrate how the application of these genomic tools to ecological model species means that we can start addressing some of the questions that have puzzled ecological geneticists for decades such as: How many genes are involved in adaptation? What types of genetic variation are responsible for adaptation? Does adaptation utilise pre-existing genetic variation or does it require new mutations to arise following an environmental change?

Next generation sequencing and ecological genetics

It is widely recognised that recent advances in DNA sequencing technology [1,2] and the development of downstream genomics tools, are changing the face of most areas of biology. For ecologists and evolutionary biologists next generation sequencing (NGS) [3] makes it more feasible than ever to identify genetic loci responsible for adaptive evolution in non-model organisms [4–6] (see Figure 1 for recent examples). In this review we will consider how the application of NGS to ecological model species is starting to provide some of the previously elusive answers to questions about the genetics of adaptation. We will not attempt to describe NGS in detail, nor give detailed descriptions of the approaches used to identify loci as these have been covered in other recent reviews, e.g. [2,4,5,7–9]. We focus mostly on ecological model species rather than classical genetic model organisms, as the former often have a well-understood ecology, including knowledge of adaptation to different environments and now, for the first time, sophisticated genetics toolkits can be developed for them. Many important questions that are fundamental to our attempts to understand the genetics of adaptation remain unanswered because we lack the necessary empirical data [10–12]. Here we consider how the discovery of adaptation

Corresponding author: Slate, J. (j.slate@sheffield.ac.uk)
* these authors contributed equally to the manuscript.

Glossary

Depth of coverage (or read depth) analysis: quantifies and compares relative number of NGS sequence reads for a given locus between ecotypes, individuals or tissues. Can be used to detect structural variation from genomic sequence [19,65]. From transcriptome sequence it can quantify gene expression levels (i.e. transcript profiling, digital transcriptomics and RNA-Seq) [66] and detect splice variants [42].

Candidate gene: studies focus on a set of genes known to be involved in a pathway affecting a phenotype. Sequencing the gene in individuals with divergent phenotypes can identify mutations, which are associated with adaptive variation.

Genome-wide association studies: (also known as association mapping or LD mapping) are an extension of QTL mapping. Statistical associations between genotype and phenotype are identified in unrelated individuals and only arise when the marker and QTL are in strong linkage disequilibrium (LD). Genome-wide association studies can map loci with greater precision than QTL mapping, as LD typically declines faster in samples of unrelated individuals.

Genome enrichment: the targeted sequencing of specific regions of the genome using NGS. For example, all the exons, large gene families or megabase-sized regions [8].

Next generation sequencing (NGS): highly parallel DNA sequencing where hundreds of thousands or millions of reads (sequences) are produced in one run. Best-known platforms are the Roche 454 FLX Titanium system, Illumina's Genome Analyser (Solexa) and ABI's SOLiD.

Population genomics: genotyping many genome-wide markers (100s–1000 s), in multiple divergent populations to identify markers that have extreme levels of differentiation ('outlier loci') and are likely to be within or close to genes involved in adaptation.

Quantitative trait locus (QTL) mapping (or linkage mapping): identification of genomic regions that explain trait variation. QTL are typically mapped by crossing individuals from different populations, generating an F_2 or backcross mapping population. Mapping panel individuals are then scored for phenotypes and genotyped at many evenly-spaced markers to test for co-segregation of markers with the focal trait. Powerful and robust for finding QTL, this approach is crude at estimating the precise location of the underlying gene(s).

RAD-tags (Restriction-site Associated DNA tags): method for typing large numbers of SNPs on the Illumina Genome Analyser. Fragments are cut by a restriction enzyme and sequenced, those fragments are over-represented in the sequence reads, and so genotypes at polymorphic sites can be reliably called. This approach differs from other SNP typing methods in that SNPs do not need to be discovered beforehand and because SNP identification and estimates of allele frequencies are obtained simultaneously, saving time and money.

Structural variation: structural polymorphisms in the genome, such as deletions, insertions, duplications, translocations and inversions that change the genome structure in a size range of kilobases–megabases. Such polymorphisms that result in a change in the number of copies of a gene or genomic region are also referred to as copy number variants (CNVs).

0169-5347/$ – see front matter doi:10.1016/j.tree.2010.09.002 Trends in Ecology and Evolution xx (2010) 1–8

genes through NGS is shedding light on three questions fundamental to evolutionary genetics: Is adaptation the result of many loci of small effect or a few loci of large effect? What type of genetic variation enables adaptation (i.e. point mutations in coding regions, regulatory changes, inversions or gene duplications)? What is the source of this adaptive genetic variation? Of course, these questions do not constitute an exhaustive list of all the important questions we need to address about adaptation genomics, but they provide a useful and testable starting point from which more complex questions might arise.

Why 'genomicise' ecological model organisms?

Ecologists have often had a good idea of the main traits involved in adaptation in their study organisms, but have lacked the tools to identify the genes underlying these adaptations. In contrast, geneticists studying classical model organisms have been able to examine the genetic architecture of phenotypic variation, but have not always been able to identify the ecological significance of this variation. The advent of NGS means it is now relatively straightforward to generate genetic toolkits for ecological model organisms, facilitating the integration of genomic and ecological data. Earlier attempts to do this involved the application of genomic resources developed in model species to close relatives of the models in the wild (e.g. *Drosophila* [13], mice [14] and *Arabidopsis* [15]). What NGS offers is the opportunity to perform genomics studies on many additional ecologically interesting species without the requirement of a closely related genetic model organism (although the latter remains useful). Developing genomics tools for ecological organisms is desirable because we can study a wider range of phenotypic traits over evolutionary timescales and in more populations than was possible previously. Through this we are likely to gain a more realistic and comparative understanding of how selection works on natural levels of genetic variation, where this genetic variation comes from and how it is maintained.

The significance of the NGS era to ecological geneticists is that a range of genomic resources such as whole genome sequences, transcriptome (which includes the part of the genome that encodes proteins) sequences and genome-wide marker panels can be generated within the scope of a three-year grant. Typically, the sequencing is outsourced to a provider, so the researcher does not require direct access to expensive equipment. What is important to highlight is that NGS has not simply made existing techniques cheaper and faster, but more importantly, it has

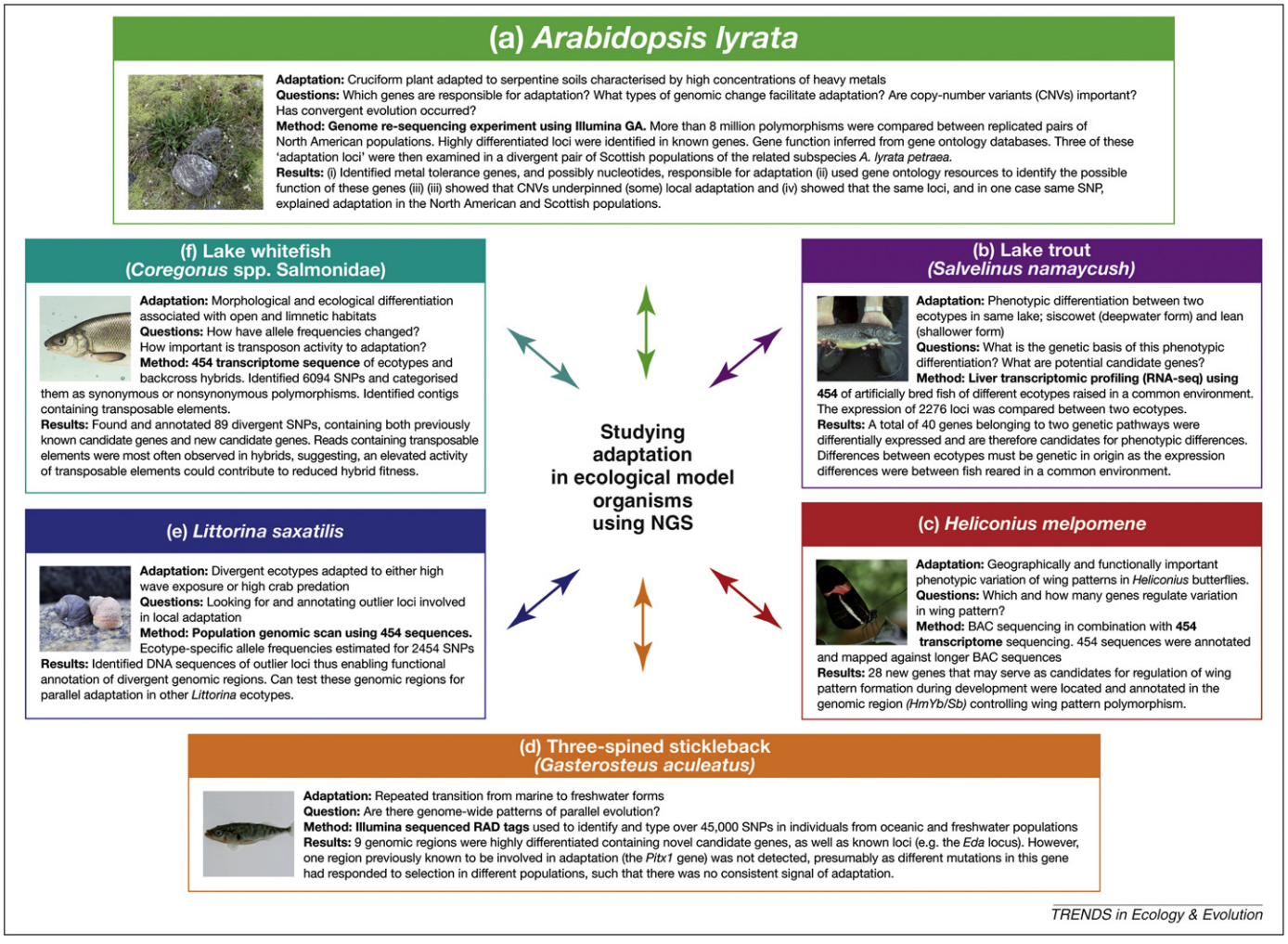

Figure 1. Examples of recent studies that used next generation sequencing technology to study adaptation in ecological model species. **(a)** *Arabidopsis lyrata* [19] (Photo: Deborah Alongi), **(b)** lake trout *Salvelinus namaycush* [67] (Photo: Ze Wrestler), **(c)** *Heliconius melpomene* [42] (Photo: Richard Bartz), **(d)** three-spined stickleback *Gasterosteus aculeatus* [29] (Photo: Piet Spaans), **(e)** *Littorina saxatilis* [16] (Photo: Juan Galindo) and **(f)** lake whitefish *Coregonus* spp. Salmonidae [41] (Photo: Ellen Edmonson and Hugh Chrisp).

enabled, for the first time, genomics studies to be conducted in any organism. For example, a single run on an Illumina GA machine, which costs $5000–$10 000 and takes around one week, will generate more data than was stored on GenBank a decade ago (16.8 Gigabases). Therefore, even from a starting point of no genetic resources in the target species and no whole genome sequence in a closely related species (referred to as a reference genome sequence), the tools required to identify genetic mechanisms involved in adaptation can be generated, e.g. [16–18]. The approaches used to identify adaptation genes with NGS data are summarised in Figure 2. It should be pointed out that although a reference genome from a related organism is not essential, when they are available, the analysis and interpretation of the data is further improved because these genomes provide valuable comparative resources for genome assembly, candidate gene discovery and subsequent analyses of sequence divergence rates and patterns [19–21].

Clearly then, the new opportunities for individual laboratories to create genomic resources for their favourite organisms, combined with the rapidly growing availability of assembled and annotated genomes in most taxonomic lineages is changing the research landscape for many evolutionary biologists and ecologists. There are still challenges associated with the analysis and interpretation of NGS data (Box 1), and so NGS should not be regarded as a simple solution to identifying genes involved in adaptation, but a great deal of progress has been made in addressing these challenges. In the remainder of this review we focus on three previously mentioned, longstanding questions in the genetics of adaptation that have been reinvigorated by NGS-based approaches. The studies we describe were conducted in non-model organisms (i.e. not traditional genetic model species like *Drosophila melanogaster*, *Mus musculus*, *Arabidopsis thaliana* and *Caenorhabditis elegans*). Although we recognise that studies of genetic model species in the wild have been and will continue to be useful in contributing empirical data to the questions outlined above, e.g. [15], here we focus on the contribution of studies in non-model organisms.

What are NGS studies revealing about adaptation?

Finding loci of small effect on phenotype

Population genetic theory has demonstrated that adaptation to new environments involves a series of genetic changes of ever smaller steps [10]. The expectation then is that a relatively large proportion of the genetic differentiation between adapted populations will involve a few genes of large effect, with the remainder explained by many loci of smaller effect. Prior to NGS being available, empirical studies used quantitative trait loci (QTL) mapping in crosses made between divergent lines, and typically found loci of large effect (reviewed in [10]). Unfortunately, these studies often lack the power to detect loci of small phenotypic effect and therefore cannot fully test theory about the genetic basis of differences between divergent phenotypes or identify all of the genes involved. The emerging complementary approach that is less biased towards finding loci of large effect is population genomics. By screening the genome for markers that have extreme levels of differentiation (outlier loci) between ecotypes, phenotypes or populations, loci with small effects can be identified when there has been sufficient time for selection to alter allele frequencies between populations. This approach has been instrumental in identifying loci in non-model species and has led to the estimate that 5–10% of the genome might be affected by natural selection (i.e. contain outlier loci) [22]. However, these estimates need to be considered cautiously. Estimates based on anonymous markers (i.e. AFLPs) are less reliable because there is no way to assess the degree of linkage, and thus the independence of loci. Furthermore, distinguishing between demography and selection as causes of genetic divergence is not straightforward [23].

How does NGS help to find loci of small effect? First, because genotyping is becoming easier and cheaper [24], it is possible to carry out mapping studies with improved power, i.e. with more individuals and more markers; the former being the main factor that limits power in QTL or genome-wide association studies [25,26]. Second, more population genomics studies will be conducted on markers that can be reliably positioned on a genome sequence, providing a more accurate estimate of the number of independent loci involved in adaptation and providing greater scope to annotate these regions and so identify potential candidate genes. Previously, AFLP-based approaches suffered from the major limitation that outliers were anonymous, such that making the advance from outlier locus to candidate genes required time-consuming steps of AFLP fragment isolation, BAC library construction and Sanger sequencing [27]. Now, because NGS can be used to carry out genome or transcriptome assembly and outlier locus detection with thousands of markers simultaneously (Figure 2), the identification of candidate genes is greatly simplified [16]. The third big advantage of NGS-based methods is that several approaches to locus discovery that used to be carried out in separate experiments, can now be integrated into one experiment, thereby combining the advantages of the different approaches (see Figure 2).

The power of NGS-based approaches to identify genes involved in adaptation was nicely demonstrated in a recent study of marine and freshwater stickleback populations. Using a new strategy for simultaneous sequencing and genotyping called RAD-tags [28], 45 000 SNPs were identified and genotyped simultaneously using Illumina sequencing [29]. Nine genomic regions, covering about 3% of the stickleback genome, were highly differentiated between the two ecotypes and these included regions containing novel candidate genes, as well as regions containing known loci of large effect (e.g. the *Eda* locus) previously identified by QTL mapping [30]. Intriguingly, the other well known gene responsible for adaptive evolution in sticklebacks, the *Pitx1* locus [31], was not within the regions of divergence, thus there was no evidence that this locus was under strong selection between ecotypes [29]. These findings demonstrate how the QTL and population genomics approaches can complement each other; the QTL approach was able to identify loci related to specific phenotypes and the population genomics approach provided evidence of the adaptive significance of the loci. In this example a reference genome was available which facilitated the positioning of markers. However, even in the

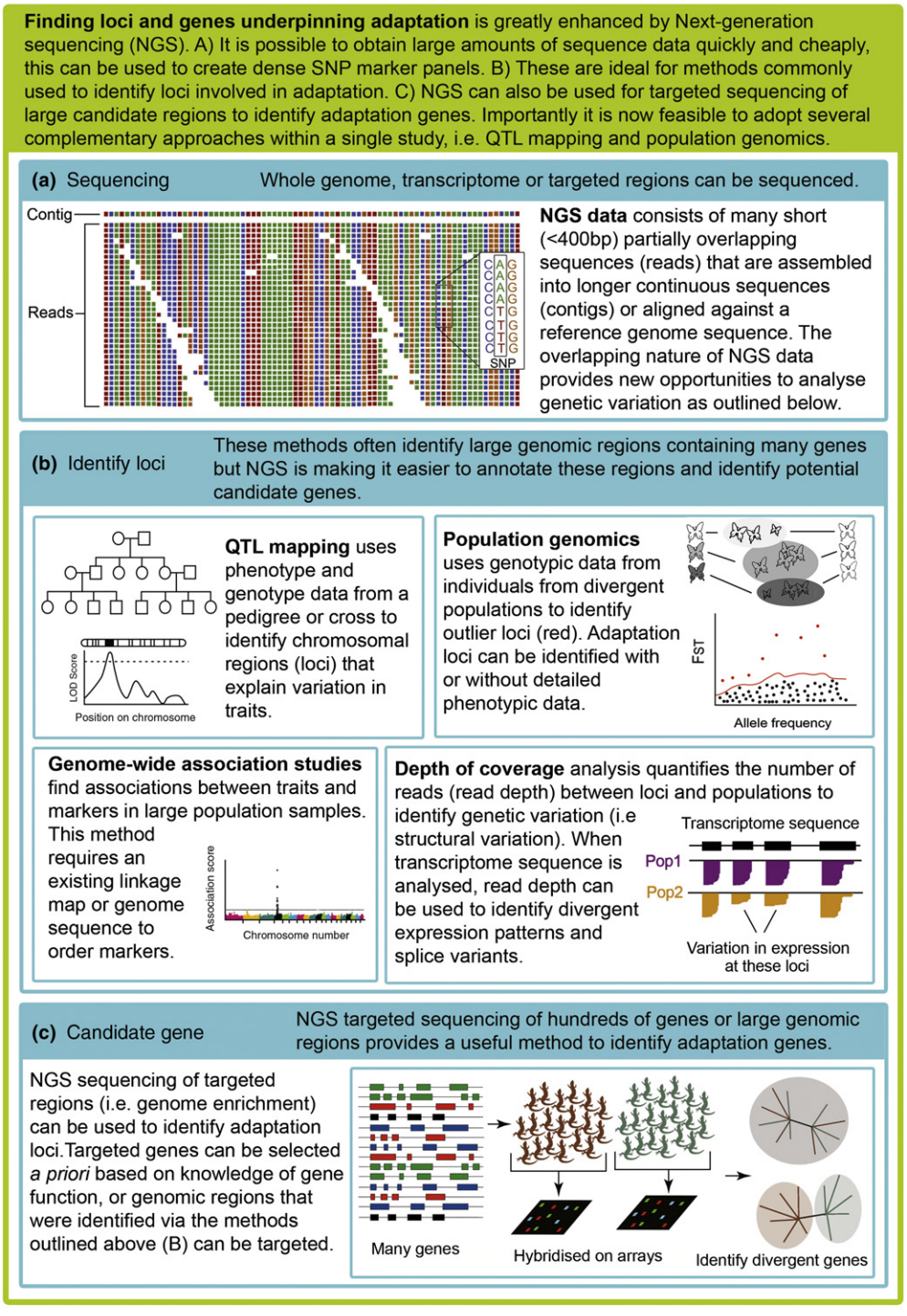

Figure 2. Summary of the common methods used to identify loci in non-model organisms; i.e. quantitative trait loci (QTL) mapping, genome-wide association studies, population genomics and candidate gene approaches; and how they have been improved and integrated by next generation sequencing.

absence of a reference genome, marker positions can be determined by building a linkage map provided that the genotyping is conducted in related individuals spanning two or more generations [29].

What are the loci of adaptation?

There has been considerable recent discussion about the type of genetic change responsible for adaptation, with most attention focusing on the relative importance of mutations that change amino acids compared to those that regulate gene expression [32,33]. Some of the first causative mutations related to adaptation that were identified involved changes in amino acid sequences (e.g. coat colour in mice [14], insecticide resistance in *Drosophila* [34], and loss of pigmentation in cavefish [35]). This is not surprising because amino-acid changing mutations are easier to identify using QTL and candidate gene approaches, the more common methods used to investigate adaptive loci in the past.

Box 1. Challenges associated with NGS

Many limitations and challenges remain with NGS. There are difficulties with storing and archiving the enormous amounts of data [55] and major challenges with data analysis and interpretation [56–58]. Below we summarise some of these issues and provide useful references that have characterised the problems and or described strategies to overcome them. We also want to emphasise that it is important not to lose sight of the ecological context of the sequence data. The ecological and phenotypic data that accompany the sequence data have to be of the highest quality for the results to be biologically meaningful.

- NGS is less accurate than Sanger sequencing and different platforms suffer from different types of errors [59–61]. Increasing depth of coverage can improve the accuracy [61].
- Mapping and assembly of NGS short reads can be difficult *de novo*, i.e. when no reference genome in a closely related species is available. Increasing the depth of coverage [58] and paired end sequencing [62] can alleviate *de novo* assembly problems.
- A big obstacle to assembly at present is informatics, i.e. limitations of hardware, software and algorithms [57,58]. A collaborative approach between ecologists, geneticists and bioinformaticians is the most practical way of analysing NGS data.
- Gene annotation and functional characterisation of sequence variation in non-model organisms remains a challenge. Functional predictions about coding sequence variation can be possible with online resources [4,63], but these are biased to model organisms and well-characterised pathways. Linking non-coding sequence variation to function is more difficult, because we have a limited understanding of the functional consequences of variation in non-coding sequences [63].
- Distinguishing a real SNP in a single gene versus genetic variation between two duplicated genes (paralogs) can be challenging and potentially will lead to false SNP discovery and typing. Longer read lengths and transcriptome profiling can help to distinguish between recent paralogs and alleles.
- Obtaining complete coverage of the transcriptome can be difficult because there is enormous disparity between expression of different genes and between different tissues. Highly expressed genes will be over-represented, while rare transcripts might be missed. Several normalisation methods are now available that can ameliorate this problem. Although it is difficult to assess the coverage of a transcriptome sequence in the absence of a reference genome sequence, it is possible to annotate the transcripts using online resources (Table S1).
- Population genomic analysis often uses pooled samples (no individual identity or barcode on each sample) to minimise sequencing costs, and then estimate allele frequencies based on read frequencies. This can be problematic because sequencing and assembly can introduce bias and errors in the data that will reduce the accuracy of allele frequency estimates. Although, analytical techniques being developed will improve these estimates [64], the impetus to pool will be reduced as sequencing becomes even cheaper.

However, convincing examples of regulatory mutations that cause adaptive divergence (e.g. pelvic reduction in threespine stickleback [36], melanism in *Drosophila* [13]) have also now been found. We do not yet know to what extent amino-acid changing and regulatory mutations influence polygenic characters, although it seems likely that the answer can depend (in part) on the evolutionary time-scale in question [37,38]. NGS provides much greater scope to identify both mutations in coding regions and regulatory changes underpinning adaptation (see Figure 2) providing empirical data that will help to resolve this debate.

More recently, studies using NGS have also demonstrated that adaptation may involve additional forms of genetic change, such as structural variation (large (kilobase-megabase) deletions, insertions, duplications and inversions) and splice variants. Identifying and typing these forms of genetic variation was difficult in non-model organisms prior to NGS and so their evolutionary significance was largely ignored. Although gene duplication (one of the forms of structural variation) has long been recognised as an important process that could generate novel genes and play a key role in adaptation, it was considered to occur at a relatively slow rate [39,40]. What is now apparent is that structural variation is more pervasive and dynamic than previously thought and that it might represent a large degree of intraspecific genetic variation [40]. A recent landmark paper on adaptation to serpentine soils in *Arabidopsis lyrata* (Figure 1) exemplifies an NGS approach [19]. Using depth of coverage analysis, copy number variants (CNVs) were identified that explained at least some of the adaptation to different soil types [19].

Genetic variation due to transposable elements and splice variants can also be uncovered by NGS-based approaches. For example, transcriptome sequencing suggests that an increased rate of transposition (the moving of transposable elements around the genome) can play a role in reduced viability of hybrids between different ecotypes of lake whitefish [41], and has also provided evidence that different splice variants might play a role in *Heliconius* butterfly wing polymorphism [42]. It is clear that NGS-based approaches such as whole genome re-sequencing, CNV-analysis and digital transcriptomics have provided the impetus to detect and quantify forms of genetic variation that were not considered in the regulatory versus coding-region debate.

What is the source of adaptive genetic variation?

Populations can adapt to new environments in two distinct ways. They can either wait for the appearance of a novel mutation, which will sweep through the population if advantageous, or alternatively, they can evolve immediately by using an allele from the standing (i.e. pre-existing) genetic variation (reviewed in [12]). Understanding which process more commonly underlies adaptation is important because much of the earliest and most influential theory on the genetics of adaptation was based on the mathematically enforced assumption that new mutations are the main source of genetic variation for adaptation (reviewed in [10]). Furthermore, the evolutionary dynamics of ancient alleles will be different to that of new mutations. Older alleles, that have been exposed to selection for generations and exist in the population at a higher frequency can reach fixation faster than young alleles, especially if their effects are recessive, as they are more likely to appear in homozygous form if older [12]. To understand whether adaptive mutations predate environmental change it is first necessary to estimate when the mutation arose, in order to compare it with the date of the environmental event that induced the adaptation (which can be estimated for example from geological events and documented anthropogenic events). As we explain below, NGS is useful here not only because is it more feasible to identify the loci to begin with, but the rapid generation of large amounts of sequencing data can provide information about a locus' history and age.

The age of adaptive loci can be estimated by sequencing adjacent regions of the genome and examining patterns of

nucleotide variation in that region. If adaptation involves new mutations that were rapidly driven to high frequencies in a new environment, then those alleles will be found on genetically impoverished haplotypes, and they will not be observed in ancestral populations or environments. Strategies for inferring the age of adaptive alleles are based around the idea of genetic hitchhiking and are summarised elsewhere [12,43,44]. Among the adaptive loci that have been identified, most seem to have been present as part of the standing genetic variation, e.g. the favoured alleles are much older than the environmental change driving adaptation (reviewed in [12]). However, there is also now at least one reasonably compelling case of an adaptation having arisen due to a new mutation that appeared after the environmental change. In this instance a mutation in the *Agouti* gene in deer mice causes a colour phenotype that confers adaptive crypsis in populations inhabiting sandhills [45].

Another source of adaptive genetic variation is admixture between two divergent populations (e.g. [46]). In the North American grey wolf (*Canis lupus*) an adaptively important coat colour polymorphism appears to have arisen through hybridisation between wild and domesticated species [47]. Colour morph frequencies differ between forested and open habitats throughout the wolf's range. Melanism in grey wolves is caused by a mutation in the *K* locus, part of the melanin synthesis pathway; a three base-pair deletion (K^B) causes the dominant inheritance of the black coat colour in grey wolves, coyotes and domestic dogs. It seems likely that the mutation arose shortly before the domestication of dogs and has reached high frequency in various dog breeds due to artificial selection. The black coat colour allele is thought to have been absent from North American and Italian grey wolf populations until relatively recently, when it was likely to have been introduced by hybridisation with domestic dogs. In North American grey wolves, black coats have reached highest frequencies in forest habitats, where it has been suggested that the melanic form has a selective advantage as it makes wolves less visible to their prey, although the latter point has not been convincingly demonstrated. In essence, the above is an example of a form of selection acting on standing genetic variation. Clearly though, the possibility that adaptive genetic variation can arise in natural populations through introgression with domesticated relatives is an intriguing area worthy of further study.

Next generation sequencing will make it easier to understand the origins and age of alleles involved in adaptation. Studies such as the deer mice and grey wolf examples (above) were time consuming because they used Sanger sequencing, which meant that target regions had to first be amplified by PCR. Now with NGS-based methods (collectively called genome enrichment [8]), targeted genomic regions can be directly sequenced without the need for amplification steps, and longer flanking regions can be sequenced as a result of the greater efficiency relative to methods that involve a PCR step. Furthermore, NGS means that it is much easier to sequence many regions of the genome at once, and therefore make comparisons between the focal region and other parts of the genome. This makes it easier to tease apart the effects of selection and demography on the genomic landscape so that the age and patterns of sequence variation at alleles of adaptive significance can be placed into context relative to the rest of the genome.

Outlook and future directions

Adaptation genomics will be performed on more organisms

Currently with NGS there is great potential to develop genomic resources for any organism in order to investigate the genetics of adaptation. For example, RADs sequenced through NGS could be used to identify and genotype thousands of SNPs in individuals from multiple populations and from an experimental pedigree, within a single round of sequencing. With these data a genetic linkage map could be built, QTL mapped and outlier loci between divergent populations identified. This could all be done in the absence of previous genetic data for the target species. It is not unreasonable to expect that such an experiment would provide evidence of loci of major phenotypic effect, outlier loci (possibly the same genes or possibly different ones with smaller phenotypic effect) that have been under divergent selection across populations and evidence of the adaptive significance of these loci in natural populations. Obtaining this detailed picture of the genetics of adaptation in non-model species was impossible for the majority of ecologists and evolutionary biologists prior to NGS, but it is now well within their reach.

Fitness effects at individual genes

Having identified loci thought to be involved in adaptation it would be desirable to bring ecological genetics studies a full-circle and measure the fitness of different genotypes at those loci. Measuring the fitness of individual loci can either be done experimentally, or by typing wild individuals of known fitness. An experimental approach was conducted using the *Eda* locus in threespine sticklebacks, and found compelling evidence that this locus can have environmentally specific pleiotropic effects on growth rate and armour plating [48]. The experiment went a long way towards explaining selection at the *Eda* locus and illustrated how differences in pleiotropy between different environments can result in the maintenance of a phenotypic polymorphism. The alternative approach to studying fitness effects at individual genes is to genotype individuals in natural populations where individual fitness has been measured. In a recent study of the gene (*Tryp1*) underlying a coat colour polymorphism in Soay sheep, fitness differences were identified between phenotypically identical but genotypically different dark sheep (i.e. those heterozygous or homozygous for dark coat-associated *Tyrp1* allele [49]). The most likely explanation for the maintenance of this coat colour polymorphism is that selection is not acting on coat colour *per se* but instead on fitness-associated genes that are in strong linkage disequilibrium with the coat colour locus.

Both these studies demonstrated a degree of genetic complexity underlying phenotypically important loci that can influence their ability to respond to selection and ultimately their adaptive scope. However, both studies examined a single locus. With NGS, a combination of a

greater capacity to identify loci, high marker density (i.e. thousands of SNPs) providing markers within (or in close linkage with) genes of interest and new SNP genotyping methods means it is feasible to study the fitness consequences of many genes at once. Studies of this kind are likely to be performed within the next few years.

Third generation sequencing

We can also expect cheaper, faster and more accurate sequencing from the next wave of sequencing technology; third generation sequencing (TGS). TGS uses very different technology to the current second generation sequencing (SGS) machines. There are four main companies developing these sequencing systems [50] and the first to become commercially available is the Single Molecular Real Time (SMRT) DNA sequencer from Pacific Biosciences, due to be released in late 2010 [51]. It is expected to produce read lengths up to 10 000 bases long and >100 000 s times faster than current NGS [51]. The longer read lengths from TGS greatly improves *de novo* genome assembly, which was a limitation of SGS data (Box 1). It is also possible to detect epigenetic changes to DNA [52] and observe RNA translation in real time [51]. Another big advantage of the longer reads will be the ability to score the phase of different alleles at linked sites (i.e. to determine which alleles are on the same chromosome), which is a major advantage when making phylogenetic inference.

Summary

At present we have a relatively short list of adaptation genes. There is little doubt that through the application of NGS to ecological model species we will see many more relevant genes being discovered and examined in detail, both at the level of the genomic landscape and in terms of their fitness in the field. Modern SNP typing and sequencing methods mean it is now possible to screen causative mutations and surrounding regions in hundreds of individuals [24,53,54], either in the wild or in controlled experiments. Perhaps the most exciting possibility raised by these recent developments is that there will be enough completed studies to test for generalisations about the rate of adaptation, the number of traits involved, the source of beneficial alleles, the magnitude of allelic effects, and how these variants are maintained as part of the standing variation within populations. By integrating the rich history of biogeography, field experimentation and long-term life history studies with cutting edge genomics tools evolutionary biologists and geneticists can test, challenge and develop new theory and greatly advance our understanding of adaptation.

Acknowledgements

We thank Susan Johnston and Stuart Dennis for help with figures. We acknowledge BBSRC, the EC, NERC and ERC for financial support. Three reviewers made constructive and insightful comments on earlier drafts of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tree.2010.09.002.

References

1 Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380
2 Mardis, E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24, 133
3 Hudson, M.E. (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Res.* 8, 3–17
4 Dalziel, A. *et al.* (2009) Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. *Mol. Ecol.* 18, 4997–5017
5 Stinchcombe, J.R. and Hoekstra, H.E. (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100, 158–170
6 Ungerer, M.C. *et al.* (2008) Ecological genomics: understanding gene and genome function in the natural environment. *Heredity* 100, 178–183
7 Marguerat, S. and Bähler, J. (2010) RNA-seq: from technology to biology. *Cell Mol. Life Sci.* 67, 569–579
8 Metzker, M.L. (2010) Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11, 31–46
9 Gilad, Y. *et al.* (2009) Characterizing natural variation using next-generation sequencing technologies. *Trends Genet.* 25, 463–471
10 Orr, H.A. (2005) The genetic theory of adaptation: A brief history. *Nat. Rev. Genet.* 6, 119–127
11 Phillips, P.C. (2007) What maintains genetic variation in natural populations? A commentary on 'The maintenance of genetic variability by mutation in a polygenic character with linked loci' by Russell Lande. *Genet. Res.* 89, 371–372
12 Barrett, R.D.H. and Schluter, D. (2008) Adaptation from standing genetic variation. *Trends Ecol. Evol.* 23, 38–44
13 Rebeiz, M. *et al.* (2009) Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* 326, 1663–1667
14 Hoekstra, H.E. *et al.* (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313, 101–104
15 Brachi, B. *et al.* (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* 6, e1000940
16 Galindo, J. *et al.* (2010) An EST-based genome scan using 454 sequencing in the marine snail *Littorina saxatilis*. *J. Evol. Biol.* 23, 2004–2016
17 Elmer, K.R. *et al.* (2010) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Mol. Ecol.* 19, 197–211
18 Emerson, K.J. *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16196–16200
19 Turner, T.L. *et al.* (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42, 260–263
20 Bonin, A. *et al.* (2009) Candidate genes revealed by a genome scan for mosquito resistance to a bacterial insecticide: sequence and gene expression variations. *BMC Genom.* 10, 551
21 Toth, A.L. *et al.* (2007) Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* 318, 441–444
22 Nosil, P. *et al.* (2009) Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* 18, 375–402
23 Excoffier, L. *et al.* (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* 103, 285–298
24 Ragoussis, J. (2009) Genotyping technologies for genetic research. *Annu. Rev. Genomics Hum. Genet.* 10, 117–133
25 Darvasi, A. and Soller, M. (1994) Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. *Theor. Appl. Genet.* 89, 351–357
26 Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*, Sinauer
27 Wood, H.M. *et al.* (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Mol. Ecol.* 17, 3123–3135
28 Baird, N.A. *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, e3376
29 Hohenlohe, P.A. *et al.* (2010) Population genomics of parallel adaptation in Threespine stickleback using sequenced RAD tags. *PLoS Gen.* 6, e1000862

30 Colosimo, P.F. *et al.* (2004) The genetic architecture of parallel armor plate reduction in Threespine sticklebacks. *PLoS Biol.* 2, e109
31 Shapiro, M.D. *et al.* (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428, 717–723
32 Hoekstra, H.E. and Coyne, J.A. (2007) The locus of evolution: Evo devo and the genetics of adaptation. *Evolution* 61, 995–1016
33 Carroll, S.B. (2008) Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* 134, 25–36
34 Ffrench-Constant, R.H. *et al.* (1993) A point mutation in a *Drosophila* GABA receptor confers insecticide resistance. *Nature* 363, 449–451
35 Protas, M.E. *et al.* (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* 38, 107–111
36 Chan, Y.F. *et al.* (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327, 302–305
37 Stern, D.L. and Orgogozo, V. (2008) The loci of evolution: How predictable is genetic evolution? *Evolution* 62, 2155–2177
38 Stern, D.L. and Orgogozo, V. (2009) Is genetic evolution predictable? *Science* 323, 746–751
39 Zhang, J. (2003) Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298
40 Korbel, J.O. *et al.* (2008) The current excitement about copy-number variation: how it relates to gene duplications and protein families. *Curr. Opin. Struct. Biol.* 18, 366–374
41 Renaut, S. *et al.* (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol. Ecol.* 19, 115–131
42 Ferguson, L. *et al.* (2010) Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the *HmYb/Sb* locus. *Mol. Ecol.* 19, 240–254
43 Sabeti, P.C. *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837
44 Slatkin, M. and Rannala, B. (2000) Estimating allele age. *Annu. Rev. Genomics Hum. Genet.* 1, 225–249
45 Linnen, C.R. *et al.* (2009) On the origin and spread of an adaptive allele in Deer mice. *Science* 325, 1095–1098
46 Rieseberg, L.H. *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301, 1211–1216
47 Anderson, T.M. *et al.* (2009) Molecular and evolutionary history of melanism in North American gray wolves. *Science* 323, 1339–1343
48 Barrett, R.D.H. *et al.* (2009) Environment specific pleiotrophy facilitates divergence at the *Ectodyplasm* locus in threespine stickleback. *Evolution* 63, 2831–2837
49 Gratten, J. *et al.* (2008) A localized negative genetic correlation constrains microevolution of coat color in wild sheep. *Science* 319, 318–320
50 Munroe, D.J. and Harris, T.J.R. (2010) Third-generation sequencing fireworks at Marco Island. *Nat. Biotechnol.* 28, 426–428
51 McCarthy, A. (2010) Third generation DNA sequencing: Pacific biosciences' single molecule real time technology. *Chem. Biol.* 17, 675–676
52 Flusberg, B.A. *et al.* (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465
53 Morin, P.A. *et al.* (2004) SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19, 208–216
54 Slate, J. *et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* 136, 97–107
55 Richter, B.G. and Sexton, D.P. (2009) Managing and analyzing next-generation sequence data. *PLoS Comp. Biol.* 5, e1000369
56 McPherson, J.D. (2009) Next-generation gap. *Nat. Methods* 6, S2–S5
57 Morozova, O. and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255–264
58 Pop, M. and Salzberg, S.L. (2008) Bioinformatics challenges of new sequencing technology. *Trends Genet.* 24, 142–149
59 Huse, S.M. *et al.* (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 8, R143
60 Wheat, C.W. (2010) Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. *Genetica* 138, 433–451
61 Harismendy, O. *et al.* (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.* 10, R32
62 Jackman, S.D. and Birol, I. (2010) Assembling genomes using short-read sequencing technology. *Genome Biol.* 11, 202–205
63 Schnoes, A.M. *et al.* (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comp. Biol.* 5, e1000605
64 Hellmann, I. *et al.* (2008) Population genetic analysis of shotgun assemblies of genomic sequences from multiple individuals. *Genome Res.* 18, 1020–1029
65 Van Tassell, C.P. *et al.* (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods* 5, 247–252
66 Wang, Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63
67 Goetz, F. *et al.* (2010) A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Mol. Ecol.* 19, 176–196