NIH Public Access
Author Manuscript
Mol Ecol. Author manuscript; available in PMC 2014 February 26.

Published in final edited form as:
Mol Ecol. 2013 June ; 22(11): 2841–2847. doi:10.1111/mec.12350.

Genotyping-by-sequencing in ecological and conservation genomics

SHAWN R. NARUM[*,†], **C. ALEX BUERKLE**[‡], **JOHN W. DAVEY**[§], **MICHAEL R. MILLER**[¶], and **PAUL A. HOHENLOHE**[**]

[*]Department of Fisheries and Wildlife, University of Idaho, Moscow, ID, USA

[†]Columbia River Inter-Tribal Fish Commission, Hagerman Fish Culture Experiment Station, Hagerman, ID, USA

[‡]Department of Botany, University of Wyoming, Laramie, WY, USA

[§]Department of Zoology, University of Cambridge, Cambridge, UK

[¶]Department of Animal Science, University of California, Davis, CA, USA

[**]Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID, USA

Keywords

genotyping-by-sequencing; next-generation sequencing; RAD-seq; SNP

The fields of ecological and conservation genetics have developed greatly in recent decades through the use of molecular markers to investigate organisms in their natural habitat and to evaluate the effect of anthropogenic disturbances. However, many of these studies have been limited to narrow regions of the genome, allowing for limited inferences but making it difficult to generalize about the organisms and their evolutionary history. Tremendous advances in sequencing technology over the last decade (i.e. next-generation sequencing; NGS) have led to the ability to sample the genome much more densely and to observe the patterns of genetic variation that result from the full range of evolutionary processes acting across the genome (Allendorf *et al.* 2010; Stapley *et al.* 2010; Li *et al.* 2012). These studies are transforming molecular ecology by making many long-standing questions much more easily accessible in almost any organism.

When studying the genetics of wild populations, it is desirable to samples tens, hundreds or even thousands of individuals. While it is now possible to sequence whole genomes for tens of individuals with small genome sizes, the sequencing of hundreds of individuals with large genomes remains prohibitively expensive, particularly where the genome sequence is unknown. Further, for the purpose of many studies, complete genomic sequence data for all individuals would be unnecessary and simply inflate the computational and bioinformatic costs. A major recent advance has been the development of genotyping-by-sequencing (GBS) approaches that allow a targeted fraction of the genome (a reduced representation library) to be sequenced with next-generation technology rather than the entire genome, even in species with little or no previous genomic information and large genomes. The

© 2013 John Wiley & Sons Ltd

Correspondence: Shawn R. Narum, Fax: 208-837-6047; nars@critfc.org.

All authors are actively involved in the field of ecological and conservation genomics and have developed expertise in genotyping-by-sequencing methods for studies in this area. The authors were co-editors of the special issue on this subject.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

subset of the genome to be sequenced in these GBS approaches may be targeted using restriction enzymes or capture probes or by sequencing the transcriptome (reviewed in Davey *et al.* 2011). In the future, as sequencing technology and computational and bioinformatic methods develop further, whole-genome resequencing may become the predominant method for ecological and conservation genomics. Currently, reduced representation approaches offer the ability to not only discover genetic variants such as SNPs but also genotype individuals at these newly discovered loci in the same data.

This special issue on 'Genotyping-by-Sequencing in Ecological and Conservation Genomics' represents a diverse set of empirical and theoretical studies that demonstrate both the utility and some of the challenges of GBS in ecological and conservation genomics. The empirical studies include demonstrations of the utility of GBS for population genomics and association mapping, as well as the development of genomic resources (i.e. large SNP data sets) for target species. The studies also illustrate some of the differences between GBS methods, in particular, aligning paired-end reads to achieve longer consensus sequences in contrast to single-end reads with shorter alignments, and double-digest versus sonication methods to fragment DNA. In addition, several papers describe advanced data pipelines for handling GBS-related sequence data and critically evaluate best practices for GBS methods and potential biases and novel features associated with GBS data. Overall, this compilation of papers emphasizes that GBS has been quickly adopted by the scientific community and is expected to become a common tool for studies in molecular ecology.

Population genomics

Genotyping-by-sequencing methods offer major advantages for population genomics by screening thousands of polymorphisms throughout the genome that are subject to the full range of evolutionary histories (variation in drift, selection, recombination, mutation) and consequences for genetic variation. Historically, most studies in ecological and conservation genetics have relied upon a small number of putatively neutral molecular markers (e.g. allozymes, microsatellites, AFLPs), covering a very limited subset of the genome. These data sets could be used to address questions related to demographic factors that affect the entire genome (e.g. diversity, gene flow and drift, effective population sizes and genetic relationships of populations), but they had limited ability to investigate specific loci that have been subject to selection and adaptive evolution. However, GBS enables researchers to identify specific genomic regions that may have experienced natural selection, in addition to improving the precision of demographic inferences by greatly increasing the number of putatively neutral markers assayed. For example, neutral markers alone may not identify distinct populations that have evolved to become resistant to specific pathogens (Bonneaud *et al.* 2011) or locally adapted to their habitat (Storz *et al.* 2009; Narum *et al.* 2010). Conversely, neutral markers may identify significant differentiation among populations based on limited gene flow or drift, but genomic regions under selection may indicate adaptive similarity that may have been either retained after isolation (Parchman *et al.* 2013) or evolved in parallel following colonization of new habitats (e.g. Hohenlohe *et al.* 2010).

Several studies in this issue utilize genome scans to search for potentially adaptive genetic variation in a population genomics context as well as estimate demographic parameters (Table 1). Included are various species of plants, marine invertebrates, marine and freshwater fish, and small mammals, making novel inferences regarding selection in natural populations in addition to measuring demographic parameters using neutral markers (Catchen *et al.* 2013b; Corander *et al.* 2013; De Wit & Palumbi 2013; Hess *et al.* 2013; Hyma & Fay 2013; Keller *et al.* 2013; Reitzel *et al.* 2013; Roda *et al.* 2013; White *et al.* 2013). Multiple papers demonstrate the utility of GBS for phylogenetic reconstruction across species (Jones *et al.* 2013; Keller *et al.* 2013; Ogden *et al.* 2013; Roda *et al.* 2013).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Additionally, three papers take advantage of GBS to identify genomic regions involved in hybridization (Hohenlohe *et al.* 2013), speciation (Jones *et al.* 2013) and divergent adaptation (Keller *et al.* 2013). Another study (Roesti *et al.* 2013) investigates stickleback populations to reveal how heterogeneous recombination rates can modulate consequences of selection and influence outlier tests for positive selection. Roesti *et al.* (2013) also use sex-specific RAD locus coverage to scrutinize sex chromosome divergence and confirm the presence of evolutionary strata in this species. All such population genomics studies face similar challenges in navigating trade-offs in sequencing effort across loci, individuals and populations. Accordingly, Buerkle & Gompert (2013) consider the question of optimizing allocation of sequencing effort in GBS between depth of coverage per locus and larger sample sizes, in order to most effectively use sequence data for population genetics.

Genome-wide association and QTL mapping studies

Screening dense markers from the genome has effectively enabled discovery of many candidate loci involved in specific phenotypic traits, either with quantitative trait loci (QTL) mapping or with genome-wide association studies (GWAS). In the last decade, these approaches have been utilized extensively in humans to identify specific genes and pathways involved human health (Hindorff *et al.* 2009) and to discover disease alleles in model organisms (Flint & Eskin 2012). As GBS does not require previous genomic information, high-density QTL mapping and GWAS studies are now being incorporated to investigate phenotypes related to biological traits in many nonmodel species in natural environments (e.g. Parchman *et al.* 2012). In this issue, Gagnaire *et al.* (2013) use RAD-seq to map phenotypic and expression QTL for ecologically relevant traits in lake whitefish (*Coregonus clupeaformis*). Additionally, RAD-seq was used in GWAS to identify regions of the genome associated with traits such as colour dimorphism in species of cichlid fishes (Takahashi *et al.* 2013), binary migration patterns in a salmonid fish (Hecht *et al.* 2013), phenotypic shell variation of land snails (*Cepaea nemoralis*; Richards *et al.* 2013) and thermal adaptation of ectothermic fish in desert streams (Narum *et al.* 2013). These studies illustrate the potential for mapping biologically relevant traits in wild populations to provide novel insight into ecological processes and to facilitate monitoring of species at risk to extinction.

Genomic resources – SNP discovery

Development of genomic resources has long been a need in the field of molecular ecology, and NGS approaches have greatly enhanced the discovery of SNPs for many nonmodel organisms (e.g. Seeb *et al.* 2011). In particular, GBS has become a highly reliable approach for identifying SNPs both within and between populations (e.g. Hohenlohe *et al.* 2011). All 21 of the empirical studies in this issue provide new SNP resources for several species, highlighting the strengths of GBS approaches for providing new polymorphisms. While GBS is clearly powerful in diploid species, two papers in this issue describe attempts to identify SNPs in polyploid species of birch (*Betula spp.*; Wang *et al.* 2013) and four species of tetraploid sturgeon (Ogden *et al.* 2013). While SNP discovery was well demonstrated in both studies, challenges remain for calling SNP genotypes for individual organisms because polyploids may have multiple copies of different alleles. Thus, further advances in SNP genotyping algorithms (e.g. Serang *et al.* 2012) are needed in order for GBS approaches to be applied for this purpose in polyploids.

Software pipelines

As next-generation sequencers can currently produce tens to hundreds of gigabases of sequence data per run (see Glenn 2011 with a recent update at http://

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

www.molecularecologist.com/next-gen-fieldguide-2013), advanced analysis pipelines have become a necessity to filter, sort and align sequence data. A pipeline for GBS must include steps to filter out poor-quality reads, classify reads by pool or individuals based on sequence barcodes, either identify loci and alleles *de novo* or align reads to an index to discover polymorphisms, and often score genotypes for each individual included in the study. The most comprehensive pipeline for handling GBS data is Stacks (Catchen *et al.* 2011), and in this issue, Catchen *et al.* (2013a) describe new features in Stacks to calculate population genomic statistics (such as F_{ST} and nucleotide diversity), create smoothed distributions using sliding window averaging across the genome and produce output genotype files specifically formatted for commonly used downstream analysis packages. Senn *et al.* (2013) describe an extension to the Stacks pipeline, using the assembly program Cortex to assemble paired-end reads at RAD loci and call SNPs in the assembled contigs. Tools for this paired-end assembly step are also explored by Davey *et al.* (2013) and Hohenlohe *et al.* (2013). These pipelines provide bioinformatics solutions for GBS studies and are broadly applicable to many species.

Addressing biases of genotyping-by-sequencing

Genotyping-by-sequencing methods using restriction enzymes (Miller *et al.* 2007; Baird *et al.* 2008; van Orsouw *et al.* 2007; Andolfatto *et al.* 2011; Elshire *et al.* 2011; Peterson *et al.* 2012; Parchman *et al.* 2012) can produce data with unique characteristics, resulting from factors such as restriction-site polymorphism or correlations of restriction fragment length with read depth. These features of GBS data and the genotyping biases they can produce are reviewed in detail by Davey *et al.* (2013), while Gautier *et al.* (2013) and Arnold *et al.* (2013) focus on the impact of restriction-site polymorphisms on population genetics estimates. Gautier *et al.* (2013) consider the effect of allele dropout on genotyping and F_{ST} calculations using both individuals and pools. Arnold *et al.* (2013) evaluate several additional population genetics statistics, demonstrate that the choice of restriction enzyme and allele dropout can have substantial effects on these estimates, and assess the double-digest RAD-seq method (Peterson *et al.* 2012) as well as standard RAD-seq. The test of double-digest RAD-seq is particularly useful as this approach should in theory avoid or reduce the bias of fragment length coverage, but Arnold *et al.* (2013) find that the effects of restriction-site polymorphism on summary statistics are more pronounced with the double-digest method.

All three papers make basic recommendations for data filtering to mitigate the most serious effects of GBS biases, while proposing more sophisticated statistical techniques for identifying and correcting biased genotypes. However, the extensive work of developing these techniques and making them sufficiently general to be applied to a wide range of species and methods remains to be done. Of the empirical papers in this special issue, all apply some type of filter to remove loci with missing genotypes to address the problem of null alleles and other potential biases identified here. While filtering out poor loci is the most common suggestion to address these biases, there are not universal filter criteria that can be applied to all studies, and thus, each of these areas must be evaluated by investigators on a case-by-case basis. As a general guideline for future analyses of GBS data sets, all empirical studies should strive to demonstrate how these potential biases were addressed.

Future needs

While the papers in this issue demonstrate the strength of GBS in ecological and conservation genomics studies, they also highlight areas where further advances are needed. This includes more advanced methods to test for and correct biases associated with GBS, new methods to confront evolutionary theory with population genomic data, additional

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

analytical tools for associating genomic variation with evolutionary processes and histories, and new approaches for visualizing vast amounts of genomic data. These areas are expected to provide better conceptual understanding of selection on organisms in their natural ecosystems, along with improved knowledge of the underlying genetic basis for specific traits related to biological processes. This knowledge will also be utilized to design effective strategies for conserving functional genetic variation to allow for future evolution. The summary information provided in Table 1 also provides a useful context to compare results of different GBS methods.

In addition to advances in theory and analytical tools for genomic data, new technical variations of GBS are expected in the near future that include complete genome typing for individuals and genotyping large numbers of individuals at selected targets that are considered to be biologically relevant. Also, the potential to combine RNA-seq and GBS approaches to identify SNPs in the transcriptome associated with patterns of gene expression offers the potential to strengthen links between genomics, transcriptomics and proteomics. Indeed, GBS has greatly expanded research opportunities in ecological and conservation genomics, and further advances are expected to open nearly endless doors of study to advance our knowledge.

Acknowledgments

We thank Tim Vines, Loren Rieseberg, Jen Gow, Hilary McNaughton and Adrienne Nye for their support during the preparation of this issue.

References

Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nature Reviews Genetics. 2010; 11:697–709.

Andolfatto PD, Davison D, Erezyilmaz TT, et al. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Research. 2011; 21:610–617.

Arnold B, Corbet-Detig RB, Hartl D, Bomblies K. RAD-seq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Molecular Ecology. 2013; 22:3179–3190.

Baird NA, Etter PD, Atwood TS, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE. 2008; 3:e3376.

Bonneaud C, Balenger S, Russell AF, Zhang J, Hill G, Edwards SV. Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird. Proceedings of the National Academy of Sciences USA. 2011; 108:7866–7871.

Buerkle CA, Gompert Z. Population genomics based on low coverage sequencing: how low should we go? Molecular Ecology. 2013; 22:3028–3035.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci *de novo* from short-read sequences. G3 (Bethesda). 2011; 1:171–182.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Molecular Ecology. 2013a; 22:3124–3140.

Catchen J, Bassham S, Wilson T, et al. The population structure and recent colonization history of Oregon threespine stickleback determined using RAD-seq. Molecular Ecology. 2013b; 22:2864–2883.

Corander J, Majander KK, Cheng L, Merilä J. High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. Molecular Ecology. 2013; 22:2931–2940.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics. 2011; 12:499–510.

Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. Special features of RAD sequencing data: implications for genotyping. Molecular Ecology. 2013; 22:3151–3164.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

De Wit P, Palumbi SR. Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. Molecular Ecology. 2013; 22:2884–2897.

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

Flint J, Eskin E. Genome-wide association studies in mice. Nature Reviews Genetics. 2012; 13:807–817.

Gagnaire P-A, Normandeau E, Pavey SA, Bernatchez L. Mapping phenotypic, expression and transmission ratio distortion QTL using RAD markers in the Lake Whitefish (*Coregonus clupeaformis*). Molecular Ecology. 2013; 22:3036–3048.

Gautier M, Gharbi K, Cezard T, et al. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Molecular Ecology. 2013; 22:3165–3178.

Glenn TC. Field guide to next generation DNA sequencers. Molecular Ecology Resources. 2011; 11:759–769.

Hecht BC, Campbell NR, Holecek DE, Narum SR. Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. Molecular Ecology. 2013; 22:3061–3076.

Hess JE, Campbell NR, Close DA, Docker MF, Narum SR. Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. Molecular Ecology. 2013; 22:2898–2916.

Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences USA. 2009; 106:9362–9367.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson E, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics. 2010; 6:e1000862.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Molecular Ecology Resources. 2011; 11 (Suppl 1):117–122.

Hohenlohe PA, Day MD, Amish SJ, et al. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Molecular Ecology. 2013; 22:3002–3013.

Hyma KE, Fay JC. Mixing of vineyard and oak-tree ecotypes of *Saccharomyces cerevisiae* in North American vineyards. Molecular Ecology. 2013; 22:2917–2930.

Jones JC, Shaohua F, Franchini P, Schartl M, Meyer A. Hybridization and speciation in Xiphophorus fish: a genome wide approach using RAD sequencing. Molecular Ecology. 2013 in press.

Keller I, Wagner CE, Greuter L, et al. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Molecular Ecology. 2013; 22:2848–2863.

Li J, Haipeng L, Jakobsson M, Sjodin P, Lascoux M. Joint analysis of demography and selection in population genetics: where do we stand and where could we go? Molecular Ecology. 2012; 21:28–44.

Miller MR, Atwood TS, Eames BF, et al. RAD marker microarrays enable rapid mapping of zebrafish mutations. Genome Biology. 2007; 8:R105.

Narum SR, Campbell NR, Kozfkay CC, Meyer KA. Adaptation of redband trout in desert and montane environments. Molecular Ecology. 2010; 19:4622–4637.

Narum SR, Campbell NR, Meyer KA, Miller MR, Hardy RW. Thermal adaptation and acclimation of ectotherms from differing aquatic climates. Molecular Ecology. 2013; 22:3090–3097.

Ogden R, Gharbi K, Mugue N, et al. Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. Molecular Ecology. 2013; 22:3112–3123.

van Orsouw NJ, Hogers RCJ, Janssen A, et al. Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS ONE. 2007; 2:e1172.

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA. Genome-wide association genetics of an adaptive trait in lodgepole pine. Molecular Ecology. 2012; 21:2991–3005.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Parchman TL, Gompert Z, Braun MJ, et al. The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular Ecology. 201310.1111/mec.12201

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

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM. Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. Molecular Ecology. 2013; 22:2953–2970.

Richards PM, Liu MM, Lowe N, Davey JW, Blaxter ML, Davison A. A search for the Cepaea super-gene: RAD-Seq derived markers bound tightly linked loci for shell colour and banding. Molecular Ecology. 2013 in press.

Roda F, Ambrose L, Walter GM, et al. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. Molecular Ecology. 2013; 22:2941–2952.

Roesti M, Moser D, Berner D. Recombination in the threespine stickleback genome – patterns and consequences. Molecular Ecology. 2013; 22:3014–3027.

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW. SNP discovery and applications of SNP genotyping in non-model organisms. Molecular Ecology Resources. 2011; 11:1–8.

Senn H, Ogden R, Cezard T, et al. Reference-free SNP discovery for the Eurasian beaver from restriction site–associated DNA paired-end data. Molecular Ecology. 2013; 22:3141–3150.

Serang O, Mollinari M, Garcia AAF. Efficient exact maximum a posteriori computation for bayesian SNP genotyping in polyploids. PLoS ONE. 2012; 7:e30906.

Stapley J, Reger J, Feulner PGD, et al. Adaptation genomics: the next generation. Trends Ecology and Evolution. 2010; 25:705–712.

Storz JF, Runck AM, Sabatino SJ, et al. Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. Proceedings of the National Academy of Sciences USA. 2009; 106:14450–14455.

Takahashi T, Sota T, Hori M. Genetic basis of male colour dimorphism in a Lake Tanganyika cichlid fish. Molecular Ecology. 2013; 22:3049–3060.

Wang N, Thomson M, Bodles WJA, et al. Genome sequence of dwarf birch (Betula nana) and cross-species RAD markers. Molecular Ecology. 2013; 22:3098–3111.

White TA, Perkins SE, Heckel G, Searle J. Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. Molecular Ecology. 2013; 22:2971–2985.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Mol Ecol. Author manuscript; available in PMC 2014 February 26.

Table 1

Data generated for contributions to this special issue using reduced representation GBS methods

Study	Organism	Method	# loci analysed	# samples	# groups	Study goals
Catchen *et al*.	Threespine stickleback (*Gasterosteus aculeatus*)	Single-end RAD-seq*	25 679	578 inds	9 pops	Phylogeography
Corrander *et al.*	Herring (*Clupea harengus*)	Single-end RAD-seq*	5 985	2 pools	2 pops	Population differentiation
Davey *et al*.	*Caenorhabditis elegans*	Paired-end RAD-seq*	24 828	24 pools	1 laboratory strain	Quantification of technical bias
DeWit & Palumbi	Red abalone (*Haliotis rufescens*)	Transcriptome sequencing	21 579	39 inds	3 pops	Population structure; identification of outlier loci
Gagnaire *et al*.	Lake whitefish (*Coregonus clupeaformis*)	Single-end RAD-seq*	3438	102 inds	1 hybrid backcross family	QTL mapping
Hecht *et al.*	Rainbow/steelhead trout (*Oncorhynchus mykiss*)	Single-end RAD-seq*	12 073	189 inds	2 pops	Genome-wide association mapping
Hess *et al.*	Pacific lamprey (*Entosphenus tridentatus*)	Single-end RAD-seq*	4439	518 inds	21 pops	Phylogeography; identification of outlier loci
Hohenlohe *et al.*	Westslope cutthroat trout (*Oncorhynchus clarkii lewisi*)	Paired-end RAD-seq*	77 141	97 inds	5 pops	Estimation of admixture
Hyma & Fay	Yeast (*Saccharomyces cerevisiae* & *S. paradoxus*)	Single-end RAD-seq*	5425 (*S.c.*); 9809 (*S.p.*)	77 inds	8 pops	Population structure
Jones *et al*.	Swordtail fish (*Xiphophorus* spp.)	Single-end double-digest RAD-seq†	149 362	139	26 species	Phylogenetic reconstruction
Keller *et al.*	Cichlid fish (*Pundamilia* spp. & *Mbipia* spp.)	Single-end RAD-seq*	10 663	50 inds	5 species	Population structure; phylogenetic reconstruction; identification of outlier loci
Narum *et al*.	Redband trout (*Oncorhynchus mykiss gairdneri*)	Single-end RAD-seq*	10 685	774 inds	2 pops + 1 F1 family	Association mapping
Ogden *et al*.	Sturgeon (*Acipenser* spp.)	Paired-end RAD-seq*	48 731	4 pools + 8 inds	4 species from 6 sites	SNP discovery; population structure
Reitzel *et al*.	Sea anemone (*Nematostella vectensis*)	Single-end RAD-seq*	4065	30 inds	4 pops	Phylogeography; identification of outlier loci
Richards *et al.*	Land snail (*Cepaea nemoralis*)	Single-end RAD-seq*	57 750	26 inds	1 laboratory cross	Linkage mapping
Roda *et al*.	Groundsel (*Senecio* spp.)	Single-end RAD-seq*	29 307	29 pools	29 pops	Phylogenetic reconstruction; identification of outlier loci
Roesti *et al.*	Threespine stickleback (*Gasterosteus aculeatus*)	Single-end RAD-seq*	1872	282 inds	1 F2 cross	Mapping of recombination rate; sex chromosome evolution
Senn *et al.*	Eurasian beaver (*Castor fiber*)	Paired-end RAD-seq*	30 201	10 inds	3	SNP discovery
Takahashi *et al*.	Cichlid fish (*Cyprichromis leptosoma*)	Single-end RAD-seq*	11 123	14 + 78 inds	F2 cross + 1 wild population	Linkage mapping

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Study	Organism	Method	# loci analysed	# samples	# groups	Study goals
Wang *et al.*	Birch (*Betula* spp.)	Single-end RAD-seq*	~43 000	15 inds	n/a	SNP discovery
White *et al.*	Bank vole (*Myodes glareolus*)	Genotyping-by-Sequencing‡	5979	281 inds	14 pops	Genetic diversity

*Baird *et al.* 2008.

†Peterson *et al*. 2012.

‡Elshire *et al*. 2011.

Abbreviations for populations = pops, individuals = inds.