Restriction-site Associated DNA (RAD) sequencing, a genotyping-by-sequencing (GBS) approach: Applications in aquaculture

Luca Bargelloni

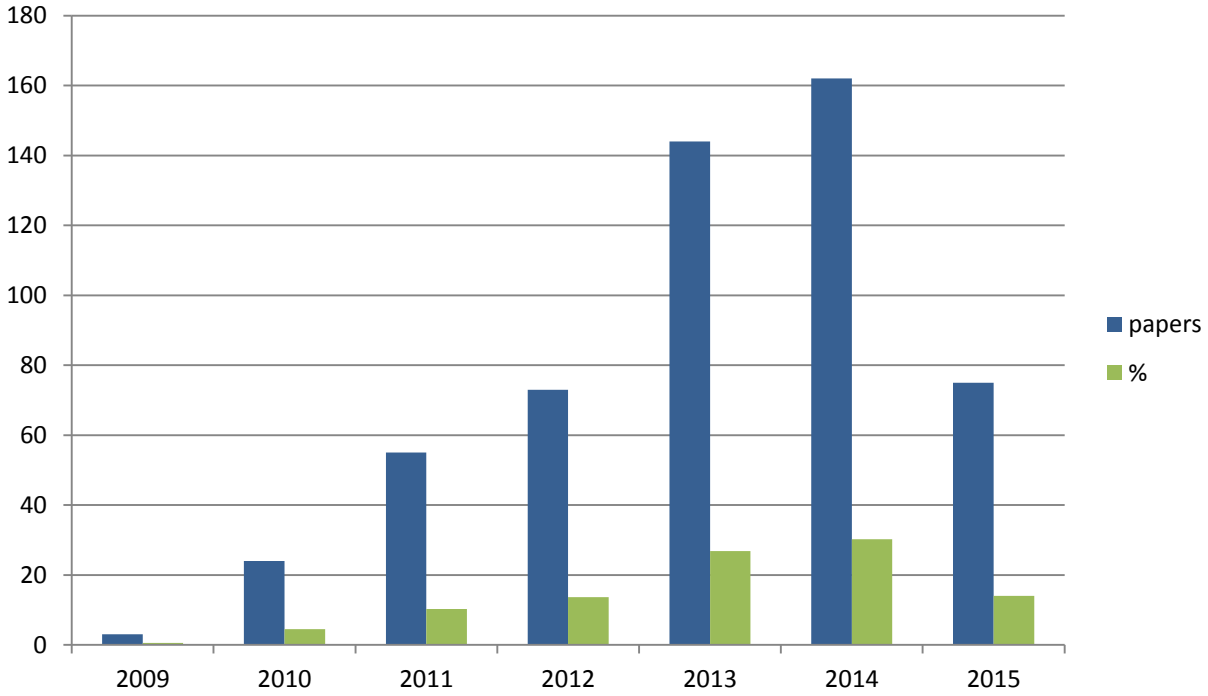
University of Padova, Italy

Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers

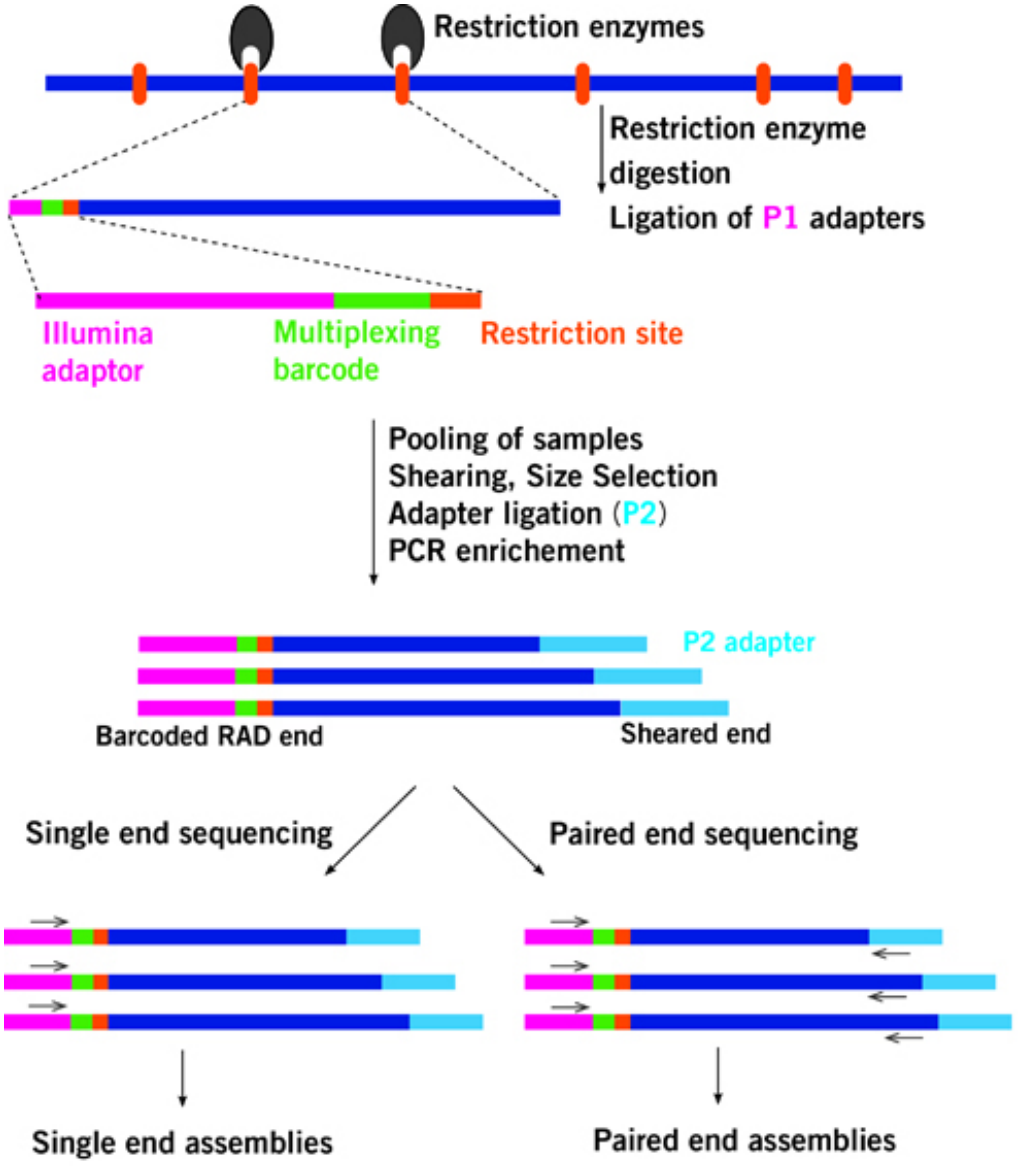
Nathan A. Baird[1☯], Paul D. Etter[1☯], Tressa S. Atwood[2], Mark C. Currey[3], Anthony L. Shiver[1], Zachary A. Lewis[1], Eric U. Selker[1], William A. Cresko[3], Eric A. Johnson[1*]

1 Institute of Molecular Biology, University of Oregon, Eugene, Oregon, United States of America, **2** Floragenex, Eugene, Oregon, United States of America, **3** The Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon, United States of America

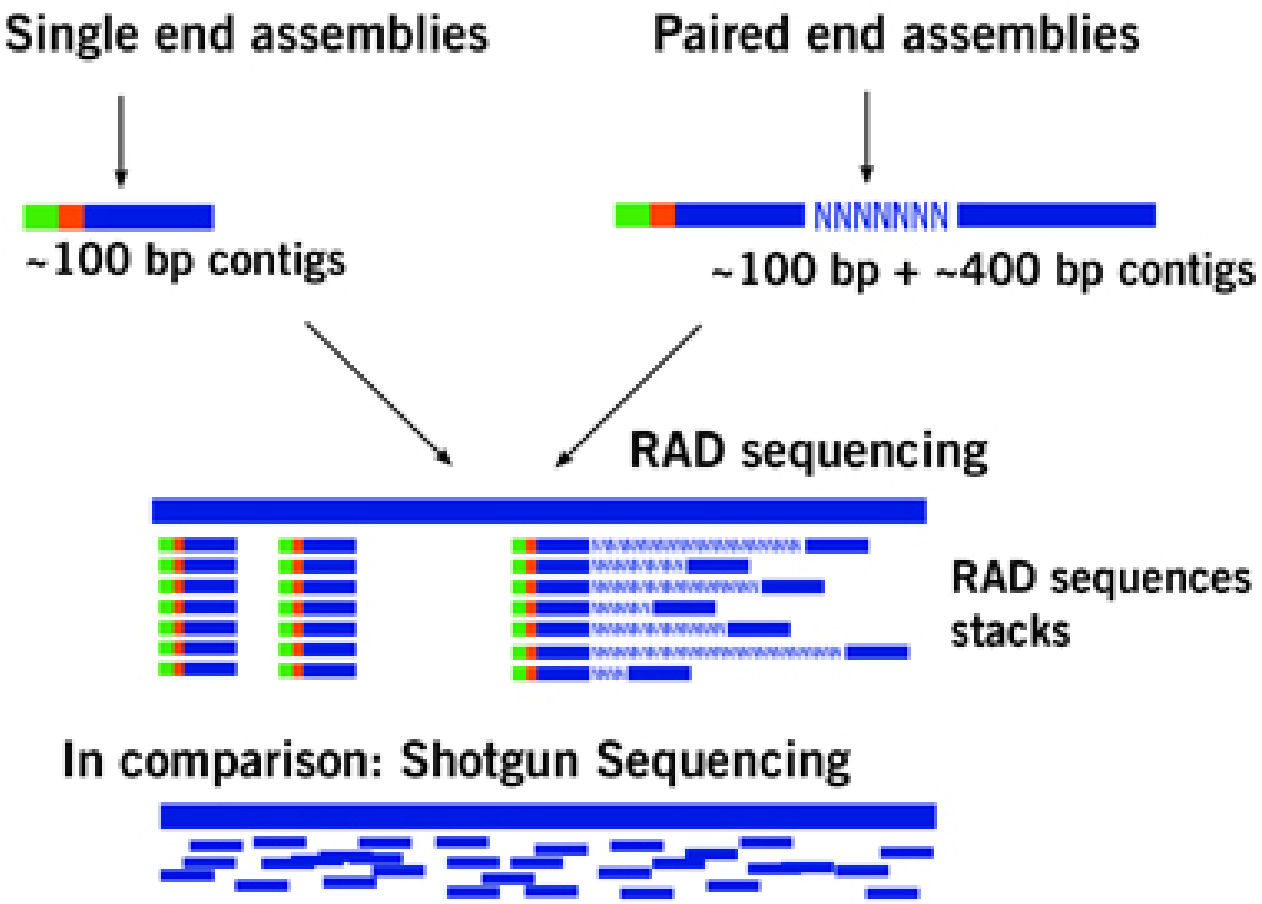
PLoS ONE | www.plosone.org 1 October 2008 | Volume 3 | Issue 10 | e3376

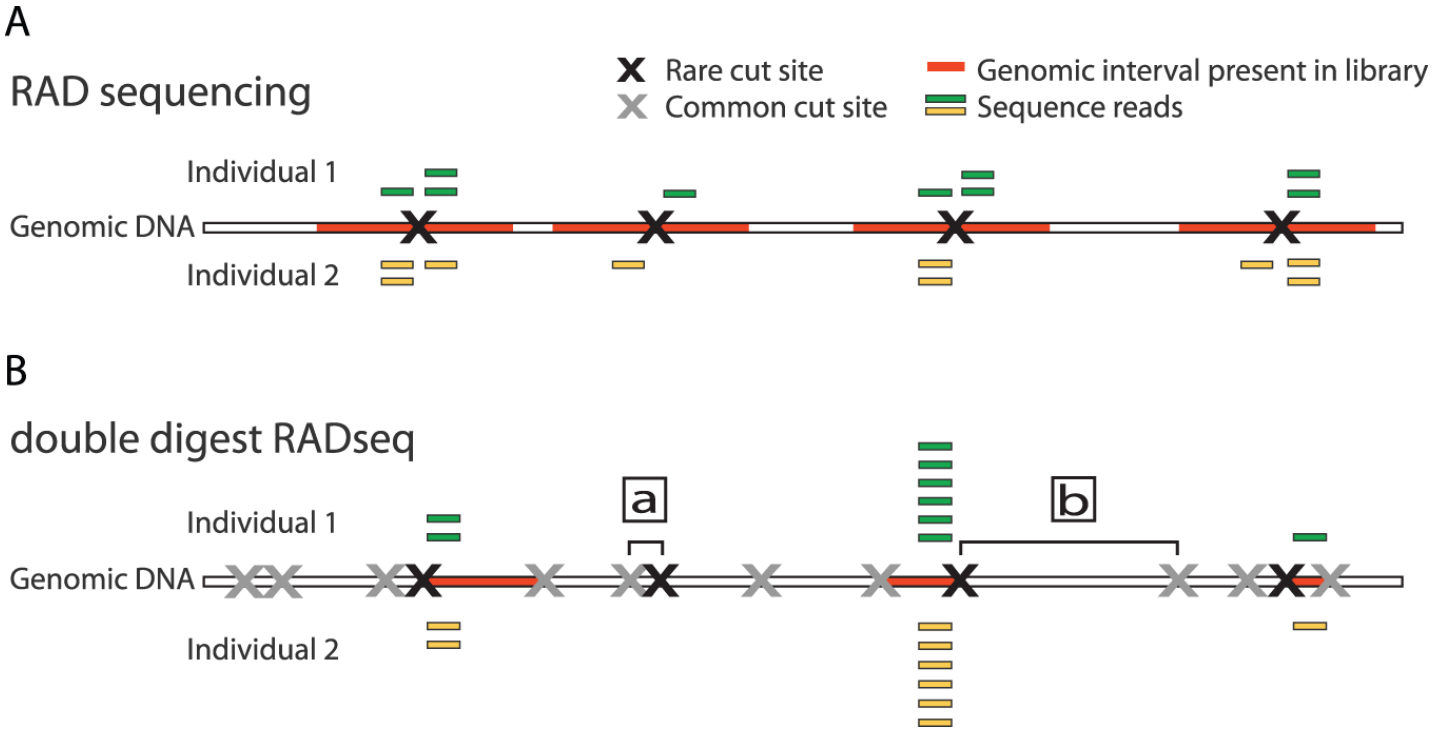


Web of Science 549 citing papers

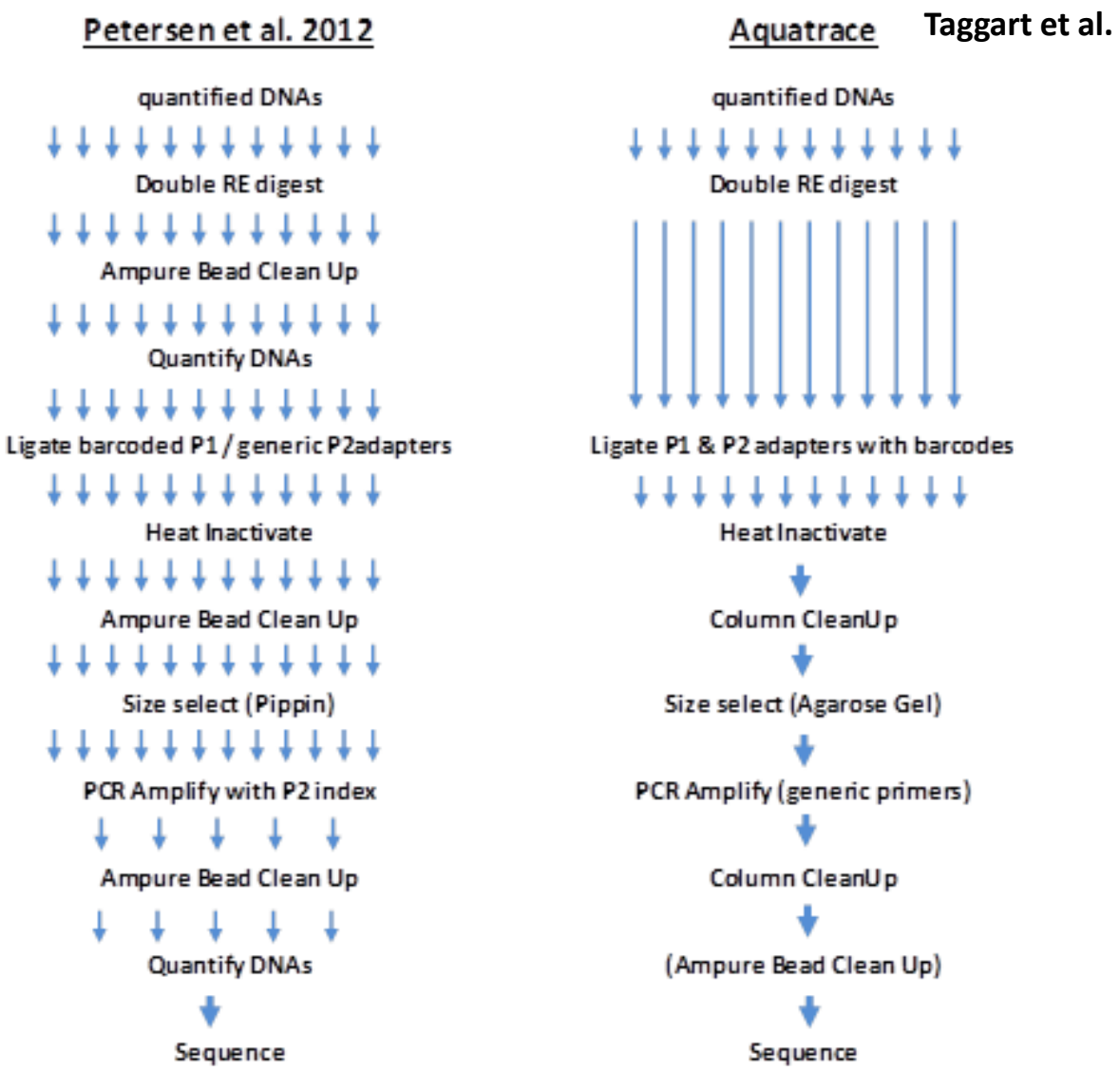


mbRAD (Miller et al. 2007, Baird et al. 2008)

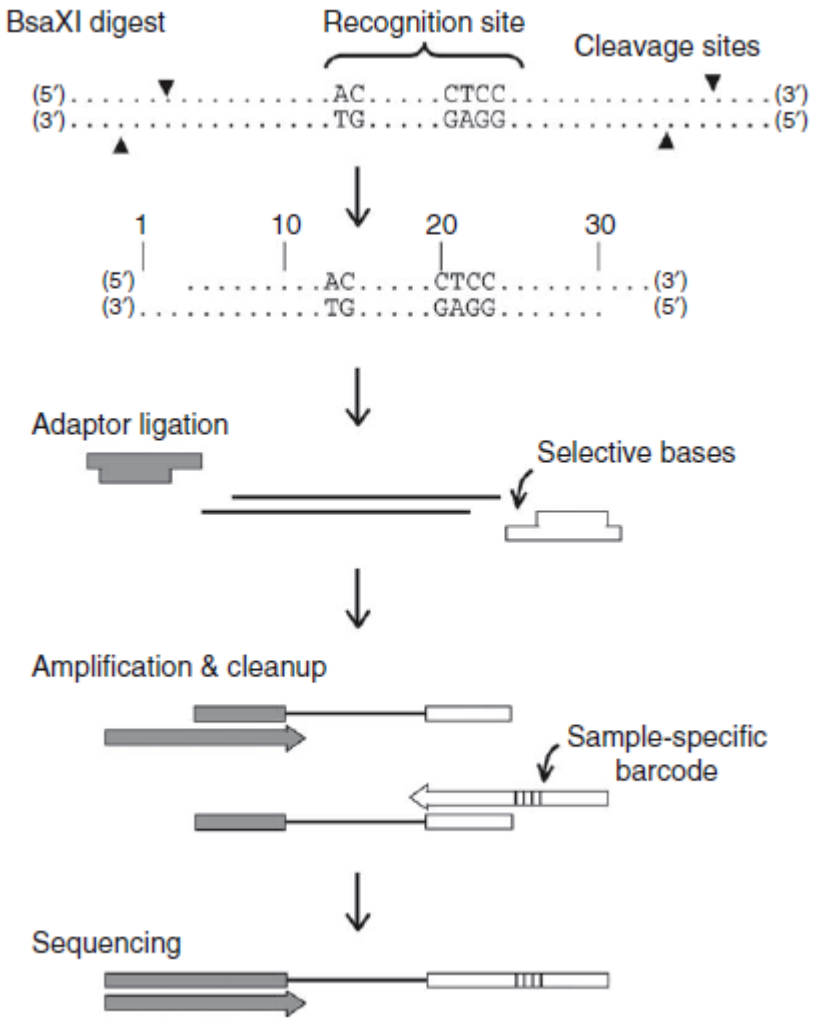

Single end assemblies
Paired end assemblies
~100 bp contigs
NNNNNNN
~100 bp + ~400 bp contigs
RAD sequencing
RAD sequences
stacks
In comparison: Shotgun Sequencing

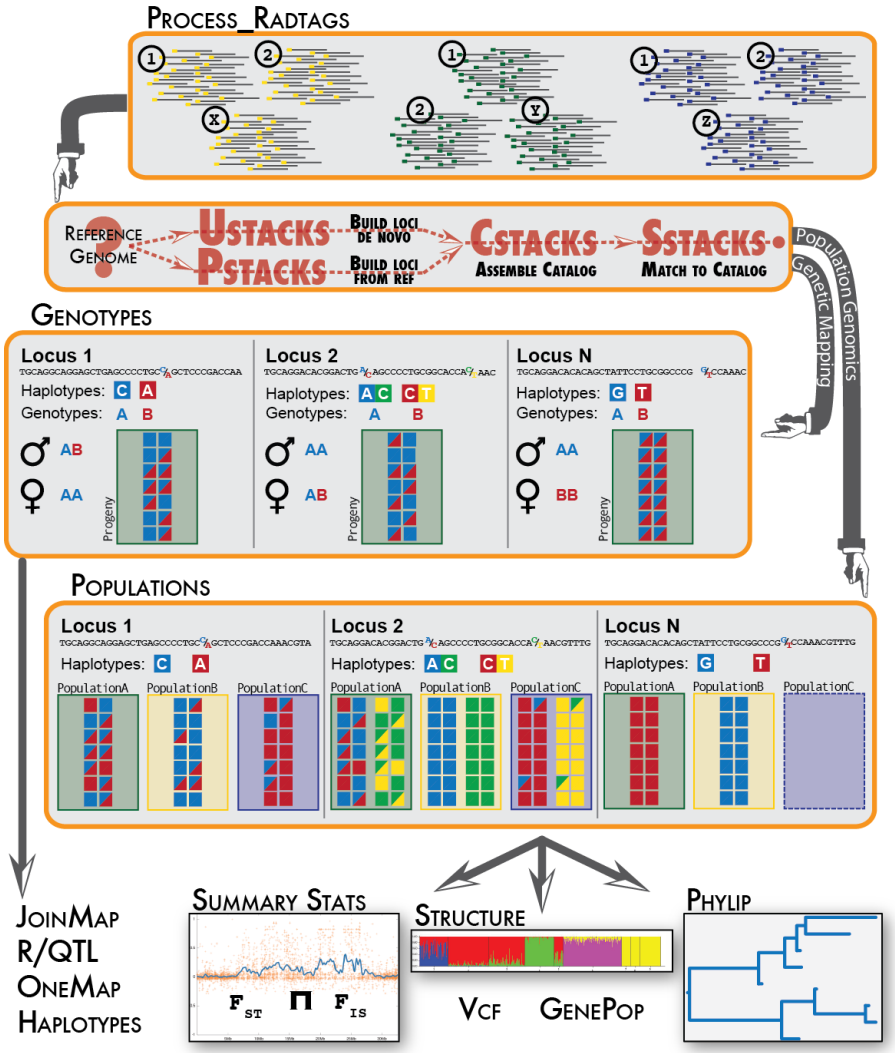


Double-digest RAD (Petersen et al. 2012)


Petersen et al. 2012
quantified DNAs
Double RE digest
Ampure Bead Clean Up
Quantify DNAs
Ligate barcoded P1 / generic P2adapters
Heat Inactivate
Ampure Bead Clean Up
Size select (Pippin)
PCR Amplify with P2 index
Ampure Bead Clean Up
Quantify DNAs
Sequence
Aquatrace
Taggart et al.
quantified DNAs
Double RE digest
Ligate P1 & P2 adapters with barcodes
Heat Inactivate
Column CleanUp
Size select (Agarose Gel)
PCR Amplify (generic primers)
Column CleanUp
(Ampure Bead Clean Up)
Sequence

2b RAD


BsaXI digest
Recognition site
Cleavage sites
(5′)
(3′)
AC
CTCC
TG
GAGG
1
10
20
30
Adaptor ligation
Selective bases
Amplification & cleanup
Sample-specific barcode
Sequencing



RAD: http://catchenlab.life.illinois.edu/stacks/
2bRAD: http://www.bio.utexas.edu/research/matz_lab/matzlab/Methods.html
2bRAD: http://people.oregonstate.edu/~meyere/tools.html
ddRAD: dDocent: https://github.com/jpuritz/dDocent

SNP discovery using RAD → SNP-chip

Table 1 Summary of the sequencing experiments for SNP discovery

	RR-Seq	RAD-Seq	RNA-Seq
Samples (number)	Farmed (40), Wild (16), Haploid (1)	Farmed (160)	Farmed (72)
Sequencing	Illumina 100 bp PE	Illumina 100 bp S&PE	Illumina 100 bp PE
Initial putative SNPs	472,072	467,268	816,570
SNPs for array design	99,097	83,151	229,754
Final SNPs on array	73,800	54,197	156,979

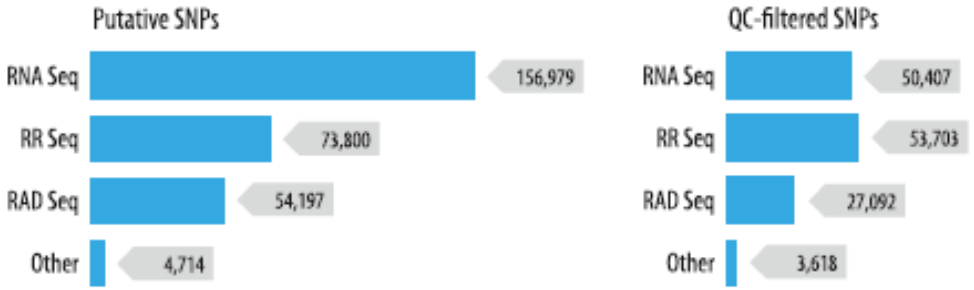


Figure 1 Source of the SNPs on the ssalar01 array. Proportion of total SNPs derived from each of the SNP discovery categories (RR-Seq, RAD-Seq, RNA-Seq and other). 'Putative SNPs' comprise the 286,021 putative SNPs placed on the array, and 'QC-filtered SNPs' comprise the 132,033 final quality-control filtered SNPs used for analysis. Note that some SNPs were detected in multiple discovery categories (see Additional file 1: Table S1).

Houston et al. BMC Genomics 2014, 15:90

RADseq → genetic linkage maps and QTL detection

Table 1 Fish samples used for ddRADseq libraries

ID	Sire strain	Dam strain	Sex ratio (total males/females)	Analysed males	Analysed females	Total analysed fish
Family 1	Red†	Clonal	64% (87/49)	54	44	98
Family 2	Red†	Clonal	72% (120/46)	36	29	65
Family 3	Red†	Clonal	93% (206/15)	28	10	38
Family 4	Red†	Clonal	92% (404/36)	43	8	51
Family 5 (28°C)	Red†	Wild*	55% (25/20)	28	22	50
Family 5 (36°C)	Red†	Wild*	96% (66/4)	66	4	70

*"Wild" refers to wild-type coloration; †"red" refers to red body colour, which is controlled by a single gene.

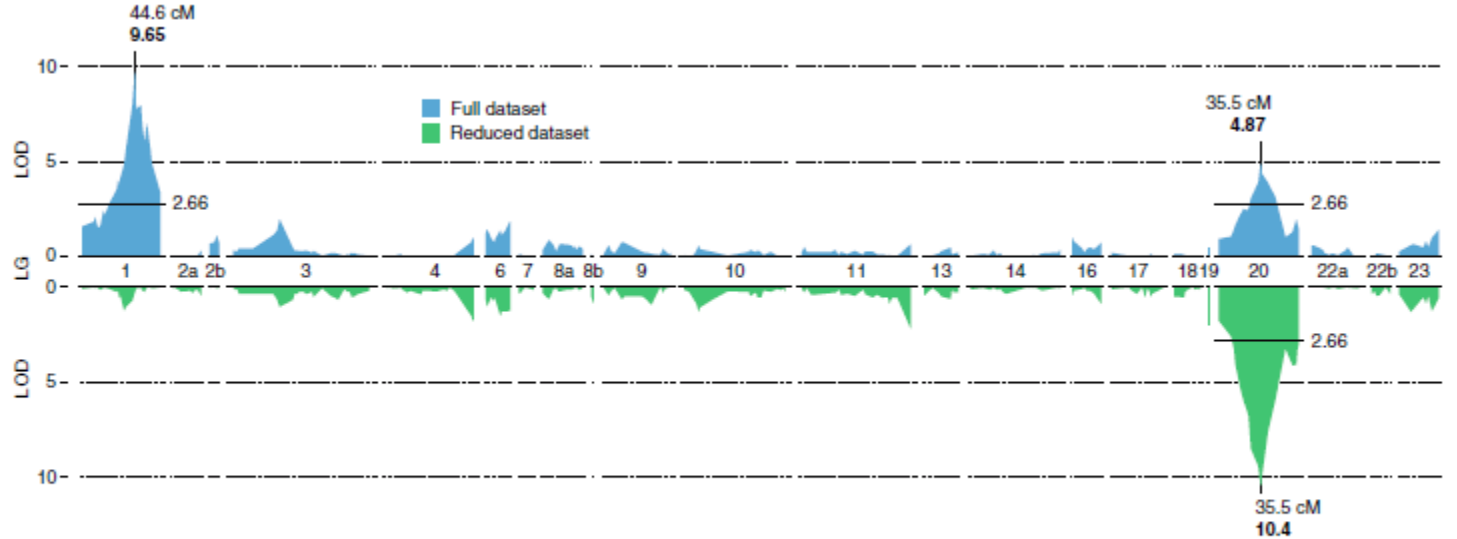


Figure 2 QTL mapping in family 1. Plots of the LOD score (sex-determining QTL search) along the linkage groups. Top graph, full dataset; bottom graph, reduced dataset (44 females and 21 males suspected to have undergone sex reversal). The numbers refer the linkage group numbers Table 3.

Palaiokostas et al. BMC Genomics (2015) 16:171

RADseq → genetic linkage maps and QTL detection

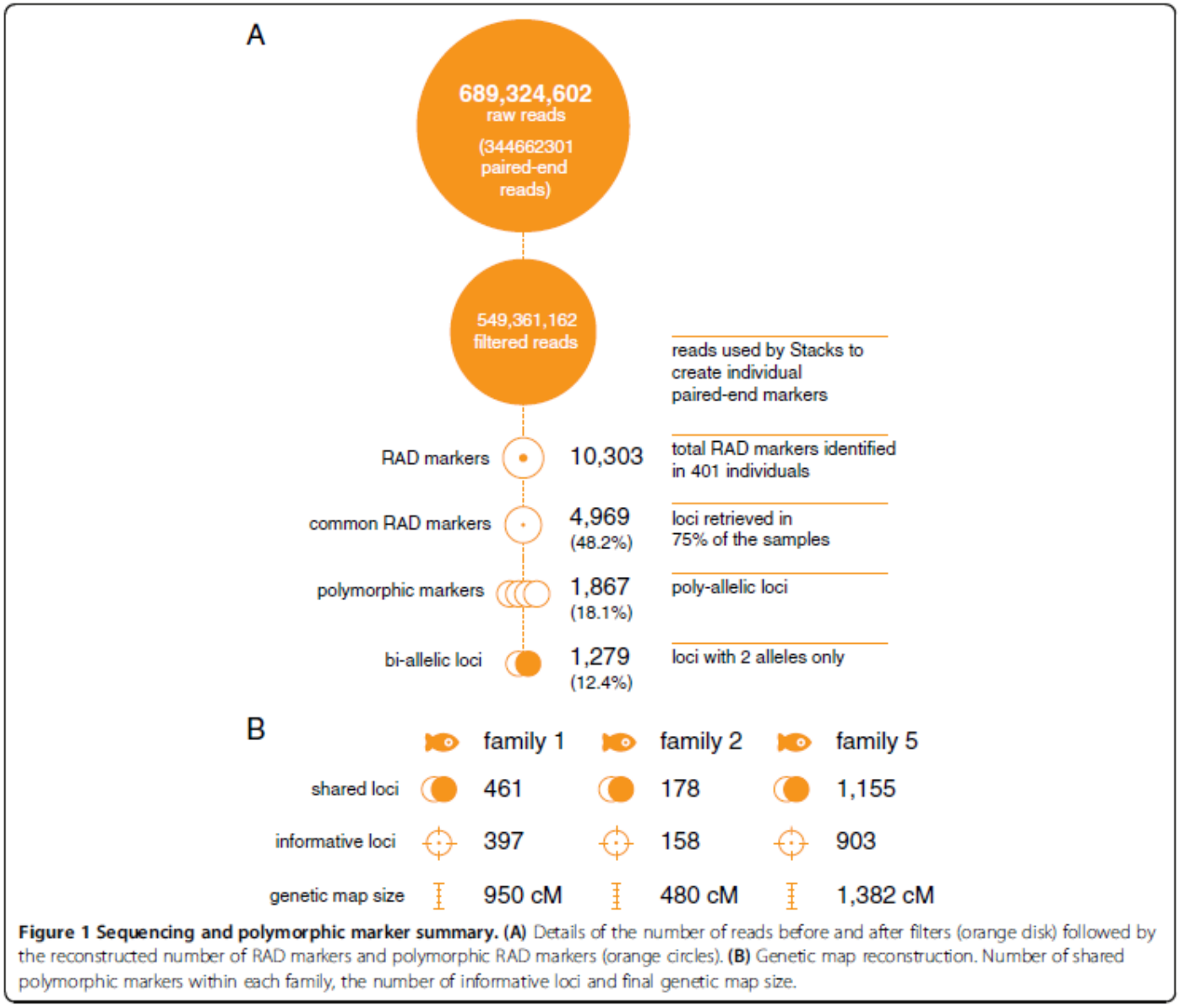


Figure 1 Sequencing and polymorphic marker summary. (A) Details of the number of reads before and after filters (orange disk) followed by the reconstructed number of RAD markers and polymorphic RAD markers (orange circles). **(B)** Genetic map reconstruction. Number of shared polymorphic markers within each family, the number of informative loci and final genetic map size.

Palaiokostas et al. BMC Genomics (2015) 16:171

Table 3. Summary of the consensus linkage map in *C. farreri*

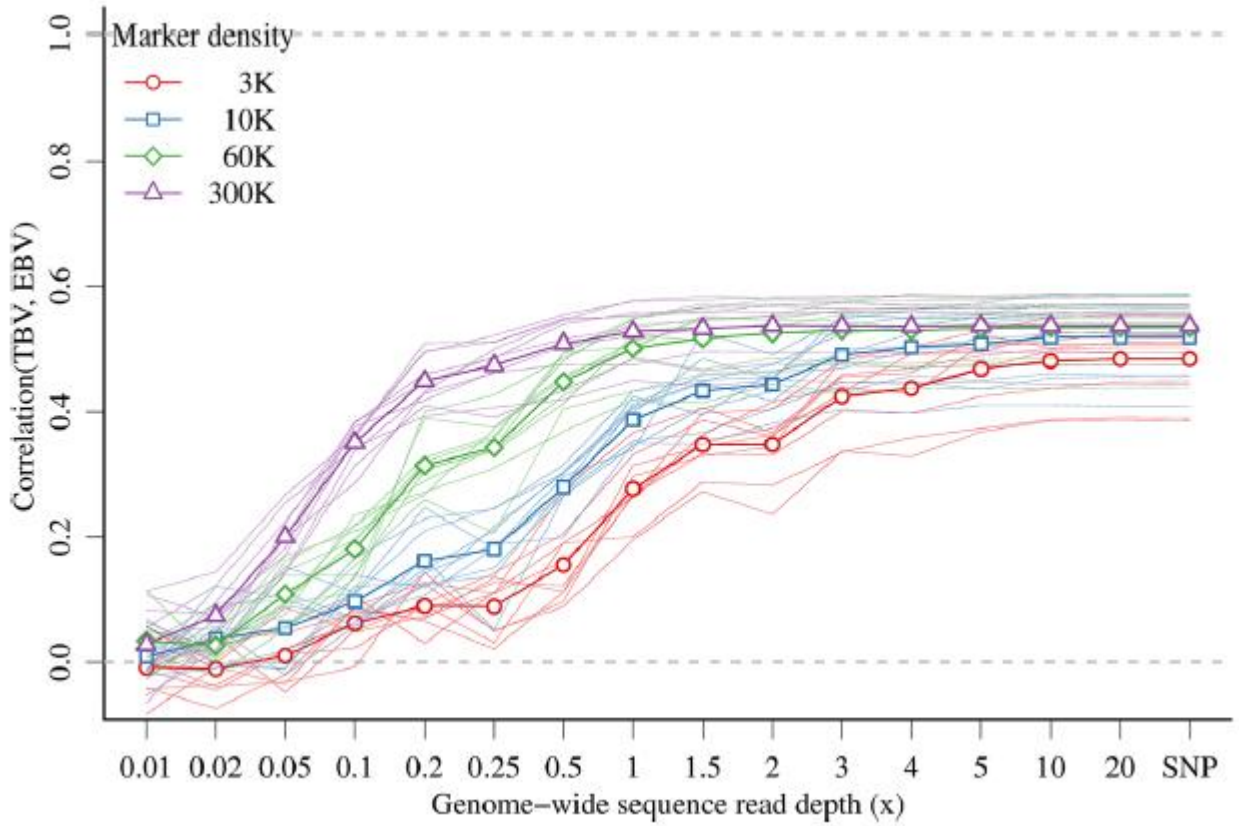
Linkage group	Number of markers	Length (cM)	Average marker interval (cM)	Female/male recombination rate	Number of annotated markers
1	274	96.09	0.35	0.97	41
2	272	88.62	0.33	0.81	43
3	271	95.36	0.35	1.04	37
4	241	75.95	0.32	1.00	39
5	235	71.48	0.31	1.37	44
6	230	99.62	0.44	1.24	53
7	223	77.77	0.35	1.00	42
8	207	68.57	0.33	0.70	35
9	203	91.94	0.46	1.30	31
10	199	54.99	0.28	2.25	30
11	189	96.59	0.51	0.97	28
12	183	92.82	0.51	0.87	26
13	172	81.04	0.47	1.14	32
14	167	65.84	0.40	1.34	18
15	166	78.2	0.47	0.62	33
16	162	84.01	0.52	0.94	27
17	147	81	0.55	0.82	23
18	142	64.36	0.46	1.07	21
19	123	79.11	0.65	1.00	20
All	3,806	1543.36	0.41	1.01	623

Jiao et al. DNA RESEARCH 21, 85–101, (2014)

2b-RAD based linkage map, 2174 (60%) tags could be unambiguously anchored to a draft genome assembly (scaffold N50 1.5 kb)

RADseq → genomic selection (?)

ddRAD 1-5k SNPs (10-50k)
RAD 5-100k SNPs
2bRAD 3-50K SNPs
Mean coverage depends on multiplexing/sequencing system (e.g typically <48 individuals on HiSeq2000 or 96 on HiSeq400 or NextSeq with >25-30X coverage)



Gorjanc et al. Genetics Selection Evolution (2015) 47:12

2b RAD, ddRAD, RAD

Time/complexity/reagent costs for library preparation

RAD>ddRAD>2bRAD

Sensitivity to DNA quality

RAD>ddRAD>2bRAD

Multiplexing (how many samples per sequencing lane)

ddRAD>RAD>2bRAD

Sequencing format (50bp SE, 100bp SE, 100bpPE…)

2bRAD (50bp SE), RAD/ddRAD (100-150 bp PE)

Number of tags

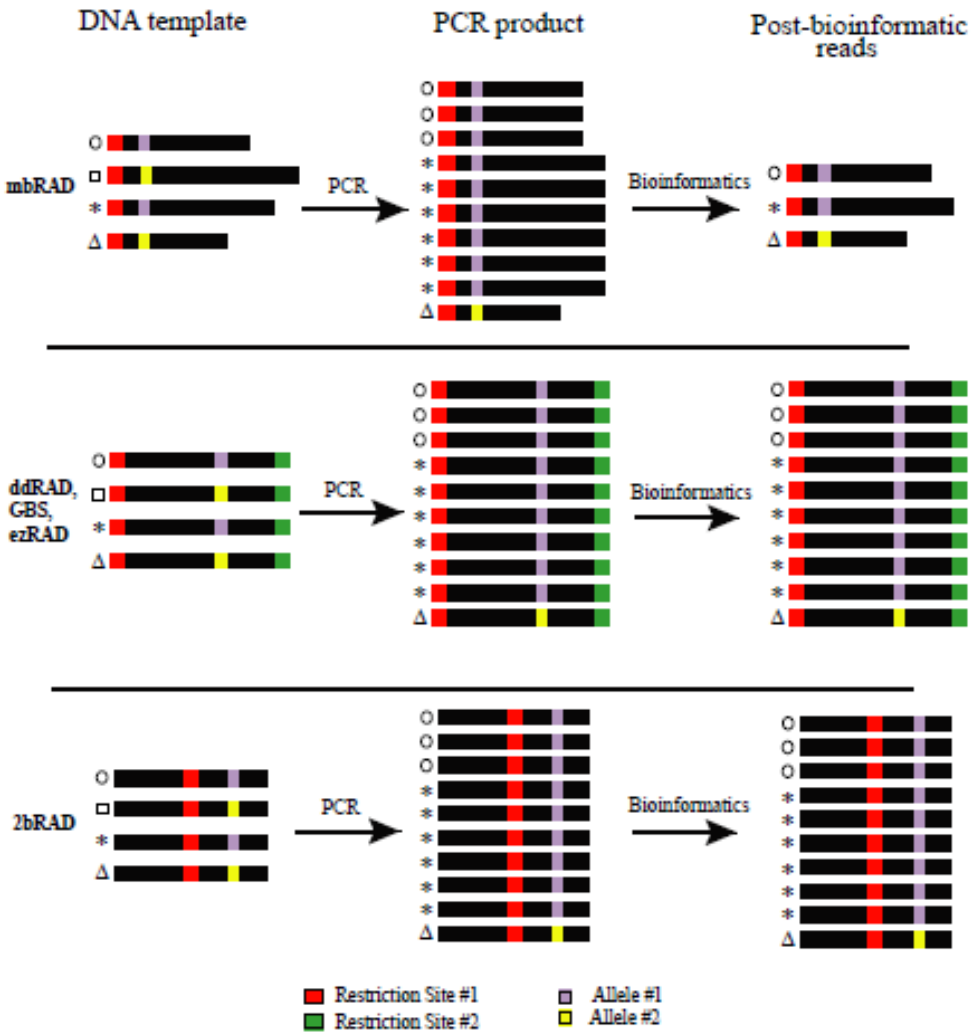
2bRAD≥RAD>ddRAD

Coverage (representation of individual libraries, representation of individual tags)

Pooled libraries (ddRAD/RAD) vs individual libraries (2bRAD)

Genotyping error (null alleles (heterozygous nt at restriction site) PCR duplicates…, bioinformatics)

2bRAD/ddRAD (PCR duplicates not detectable) as in RAD



Degenerate bases in adapters, Multiple, independent PCR reactions, DNA quality

Andrews et al. Molecular Ecology (2014) 23, 5943–5946

Table 1 Comparison of the utility, technical complexity and sources of bias for different RAD methods

	mbRAD	ddRAD	ezRAD	2bRAD
Restriction cut sites per 10 kb*	~0.2–2.4	~3.7×10^{-5}–39	~39	~2.4
Postdigest fragment reduction	Size selection	Size selection	Size selection	Selective adapters
Contigs > 200 bp†	Yes	No	Some	No
Ability to blast/annotate *de novo* contigs	High	Mid	Mid	Low
Protocol complexity (# Steps)‡	6	4	4–6	3
Level of technical difficulty	High	Mid	Low	Low
Level of technical support	Low	Low	Mid-high	Low
Insert complexity (first × bases)	Low	Low	Very low	High
PCR AT/GC content, copy number Bias among loci	Yes	Yes	Yes, No§	Yes
ID of PCR duplicates	Yes	No	No§	No¶
Uniform locus length	No	No	No	Yes
Oligos required to uniquely identify and build 96 libraries	196**	31	20–22	37
Target insert size range	200–600 bp	Customizable	Customizable	33–36 bp

Puritz et al. Molecular Ecology (2014) 23, 5937–5942
Andrews et al. Molecular Ecology (2014) 23, 5943–5946

GBS/RAD advantages

GBS can capture genetic variation that is specific to a population or family of interest

Flexibility (customizable according to targets)

Low start-up costs (new species, limited resources)

Lowering sequencing costs and increasing sequence throughput

GBS/RAD disadvantages

Labour-intensive (*e.g.* 1.5-5 days for preparing 96 samples, bioinformatics)

High sequencing costs for high sequence coverage

Genotyping uncertainty/errors (PCR duplicates, uneven coverage)

Problematic for large genomes (*e.g.* recent genome duplications, repeated regions)

Problematic in highly polymorphic species (null alleles for polymorphisms at restriction sites)