CLASE 14

Estudios de asociación genómico y selección genómica

DBT 792
GENÉTICA Y GENÓMICA EN PRODUCCIÓN ANIMAL

Profesor
Dr. José Gallardo

PLAN DE LA CLASE

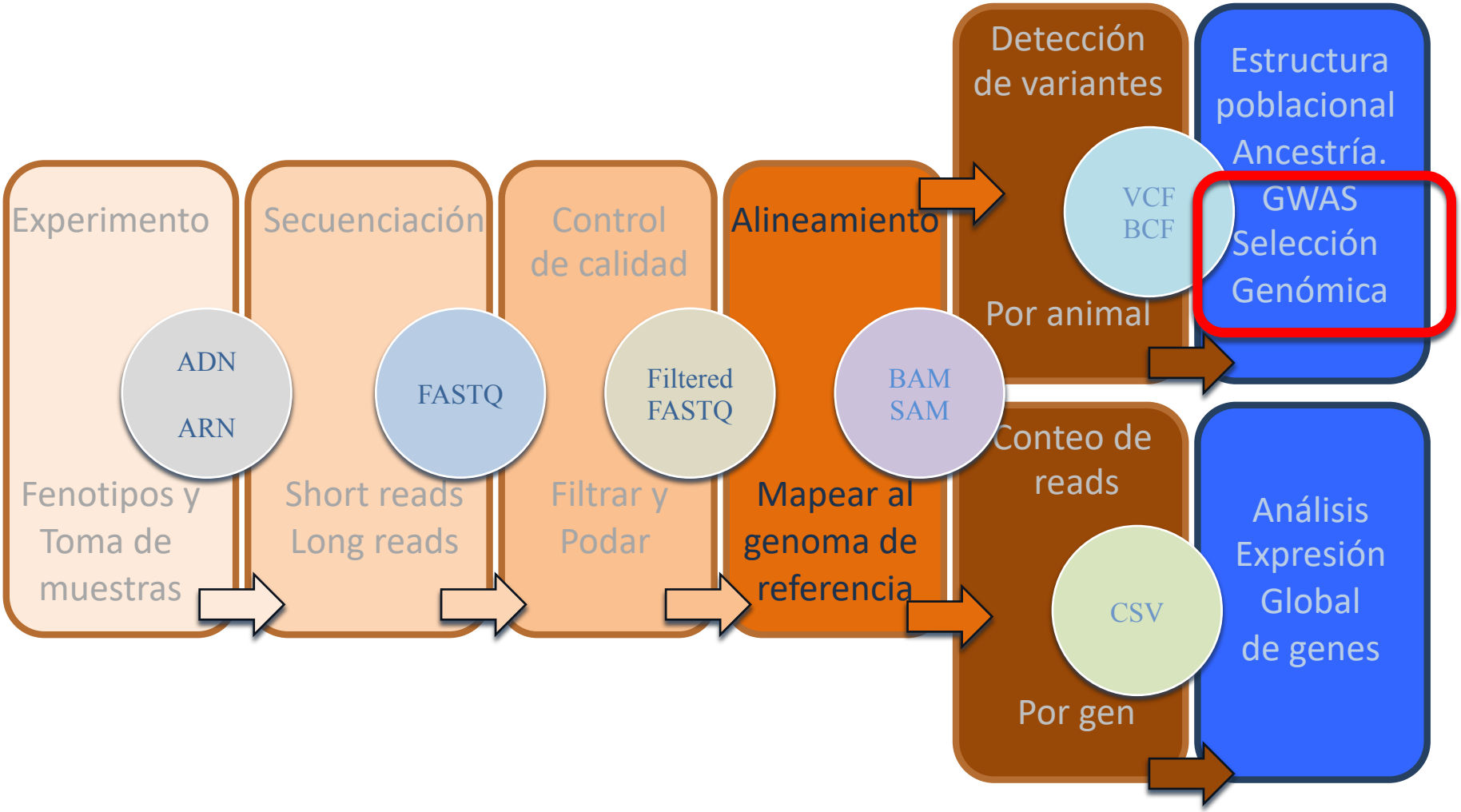
Introducción

- Repaso flujo de trabajo proyecto genómica aplicada.
- Locus de rasgos cuantitativos (QTL).
- Tipos de marcadores genéticos.
- Fundamentos de estudios de asociación genómica (GWAS) para identificar QTL.
- Fundamentos de selección genómica.
- Casos de estudio en salmón del Atlántico.

Práctica

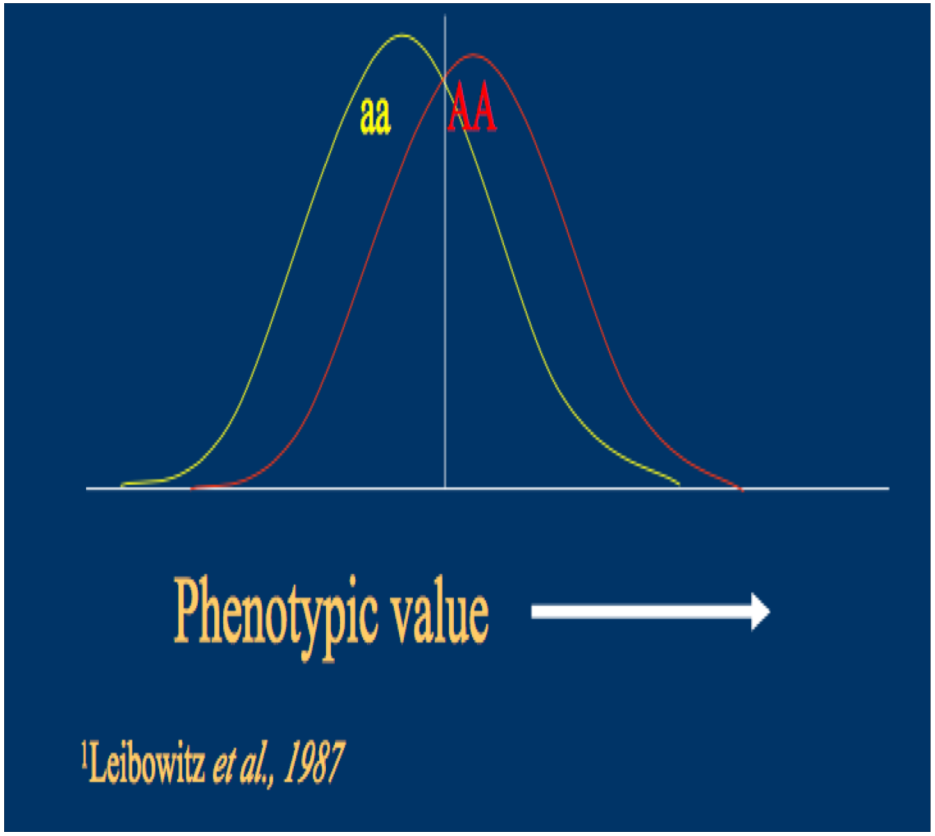
- Practica de GWAS y predicción genómica con R

FLUJO DE TRABAJO DE UN PROYECTO DE GENÓMICA APLICADA



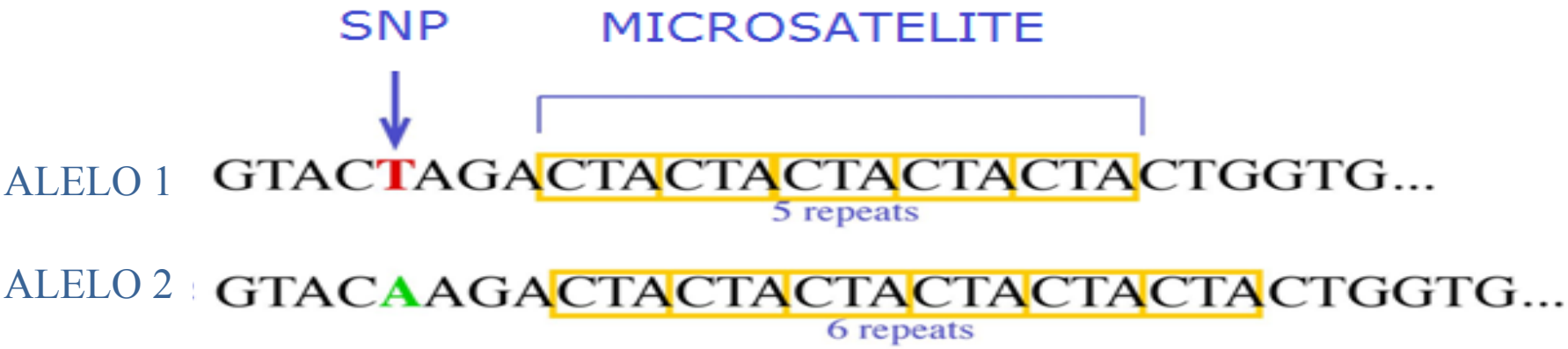
QUANTITATIVE TRAITS LOCUS (QTL)

Región o segmento de un cromosoma (locus) **cuya variación (alelos)** explica la variación (fenotipos) de un rasgo cuantitativo.



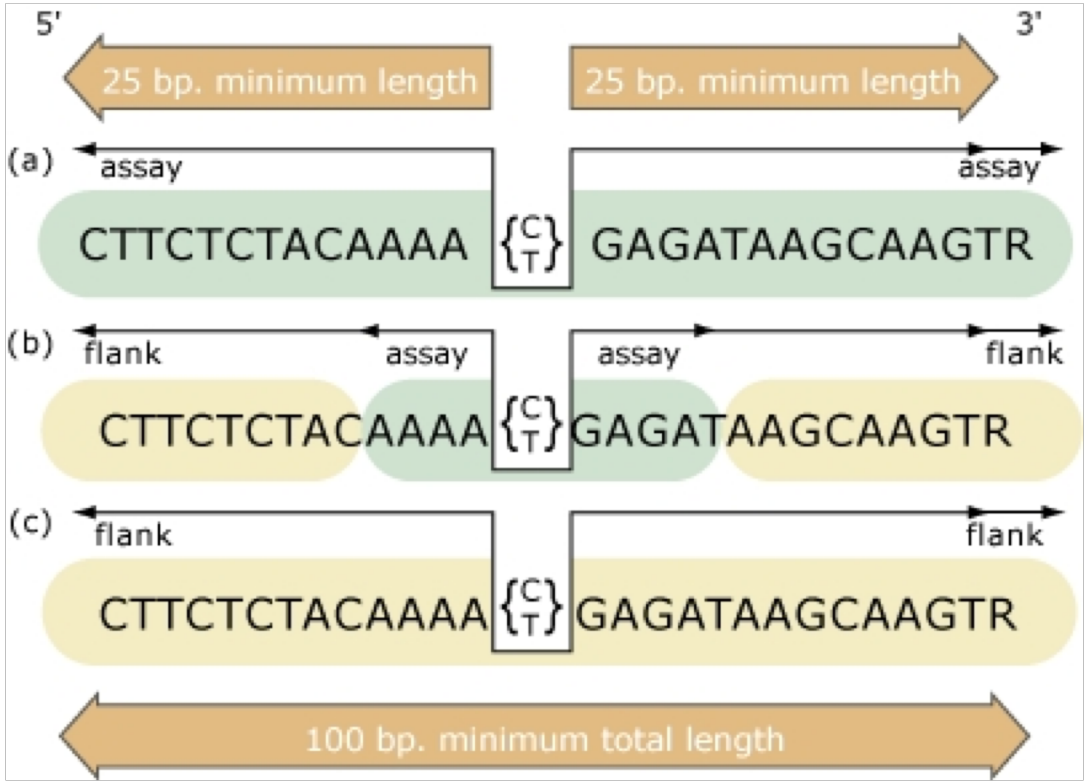
PLOS Genetics 12(11): e1006421.

MARCADORES GENÉTICOS USADOS PARA IDENTIFICAR QTL Y PARA ESTIMAR BREEDING VALUE



– Debe ser polimórfico (Genetic variations occurring in more than 1% of the population).
– No necesariamente tiene función biológica.
– Debo conocer su posición en los cromosomas.

SNP : SINGLE NUCLEOTIDE POLYMORPHISM



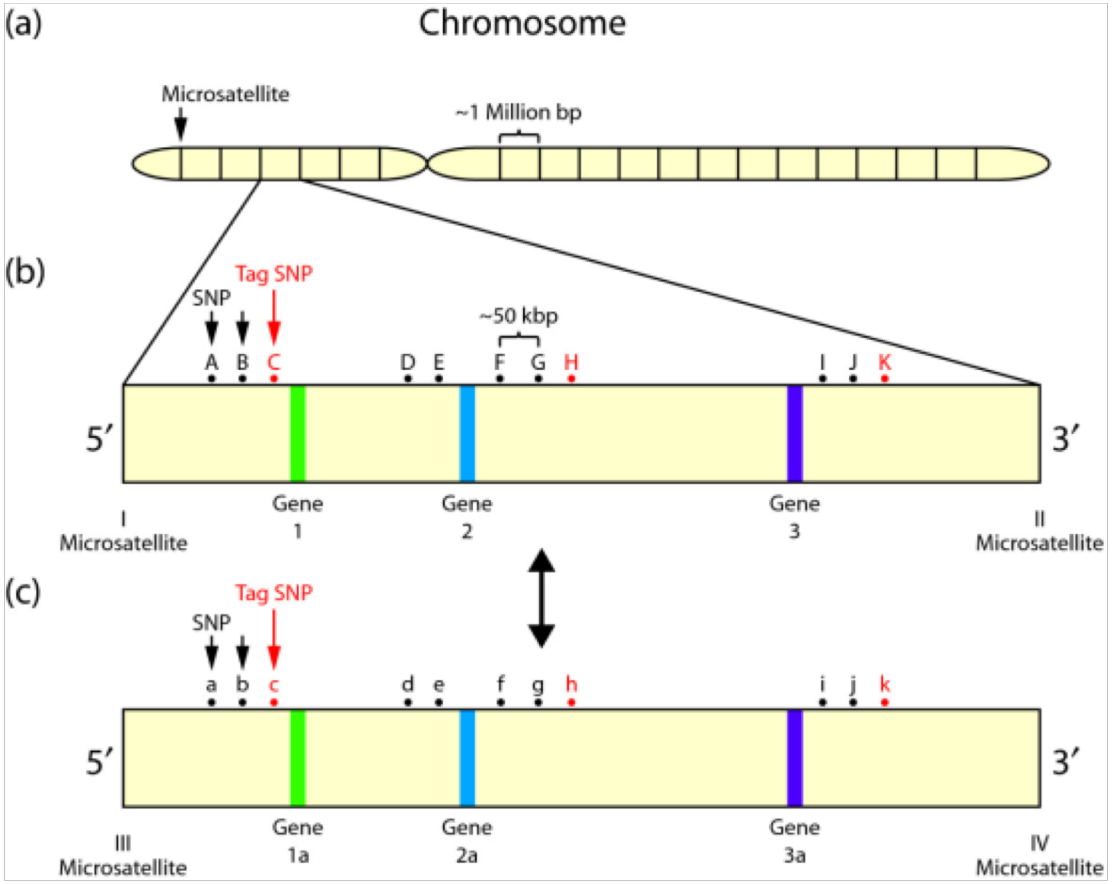
– Naturaleza bi-alélica.

2 ALELOS
C – T

3 GENOTIPOS
C/C - C/T – T/T

FUENTE: The NCBI Handbook

COMPARACIÓN MARCADORES GENÉTICOS



Microsatélites:

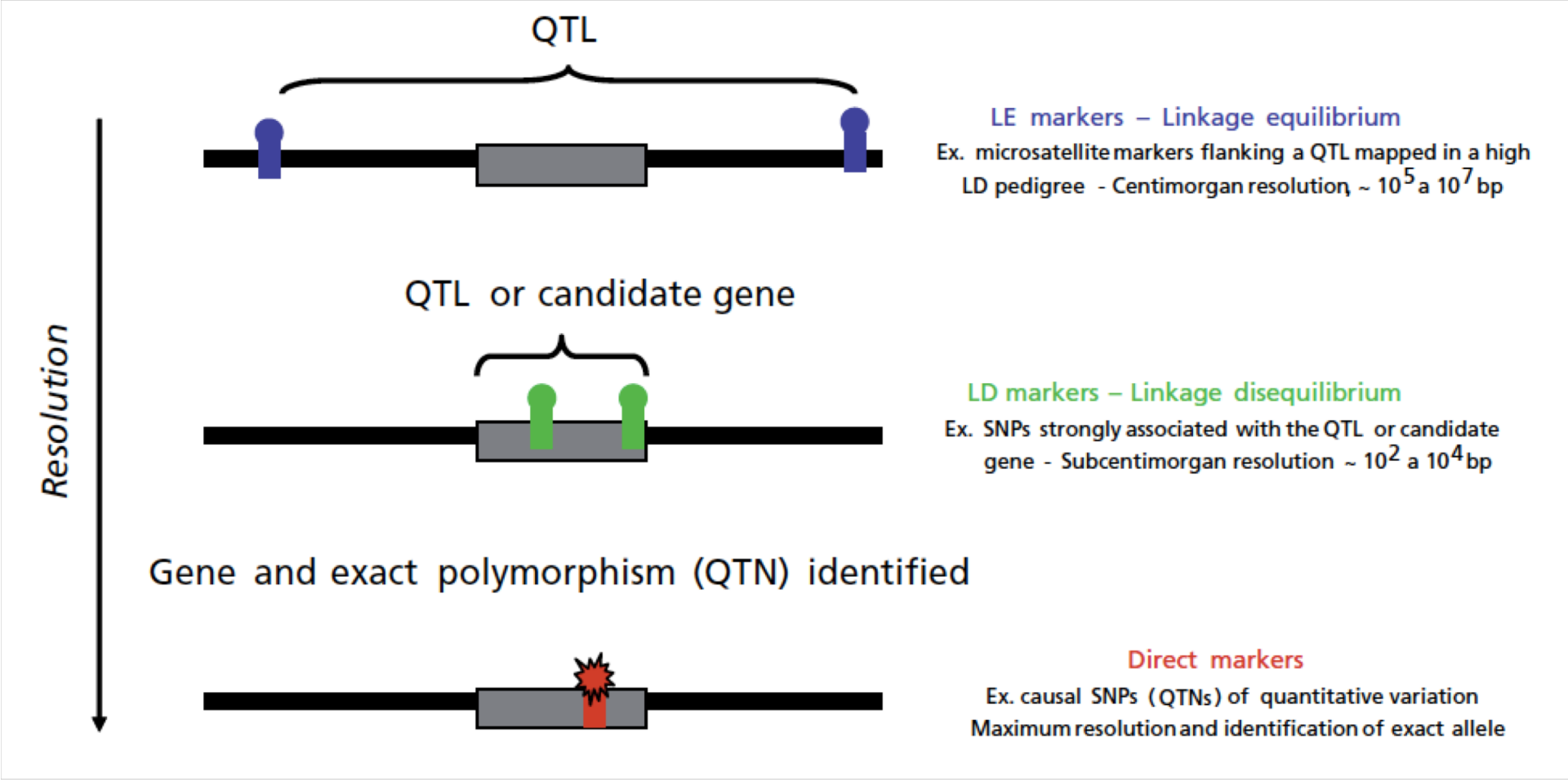
- Cientos de micros identificados en genoma de peces.
- Localizados fuera de los genes .

SNP

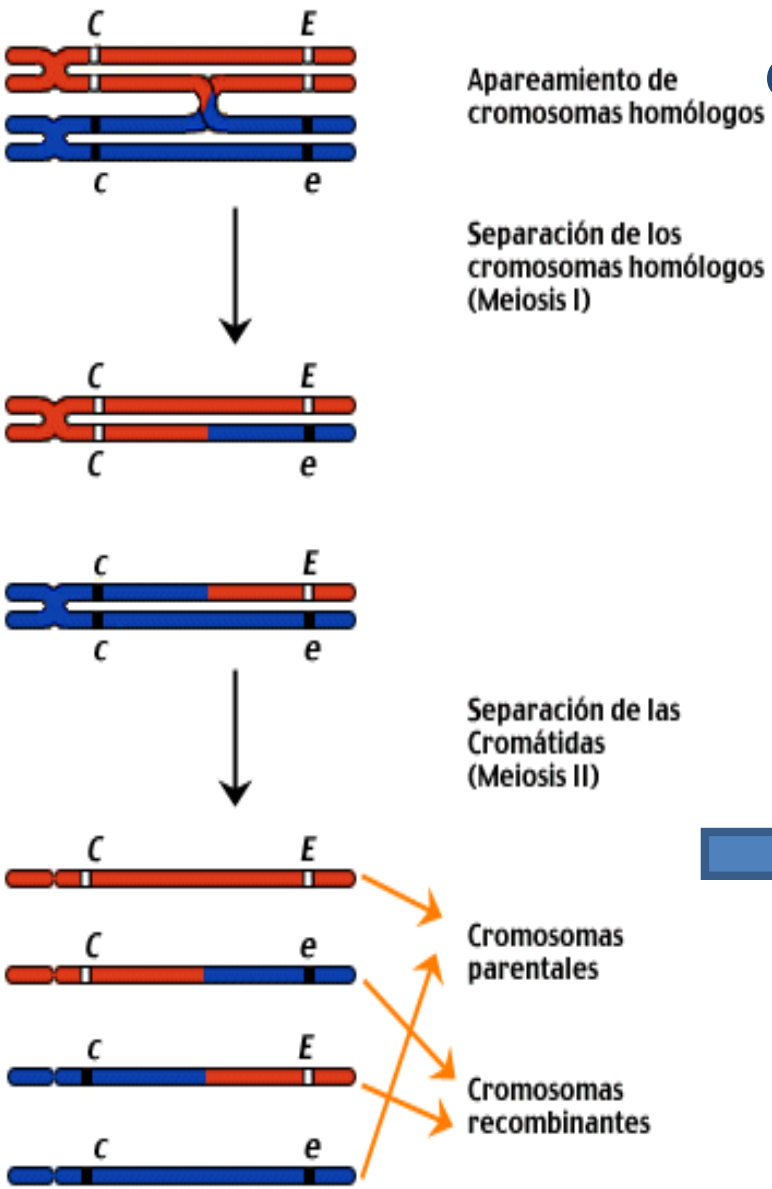
- Millones de SNPs en genomas de peces.
- Localizados cerca de los genes o dentro de ellos.

Referencia: Damani and Topol *Genome Medicine* 2009 **1**:54

CONCEPTOS BÁSICOS



MEIOSIS - EQUILIBRIO DE LIGAMIENTO



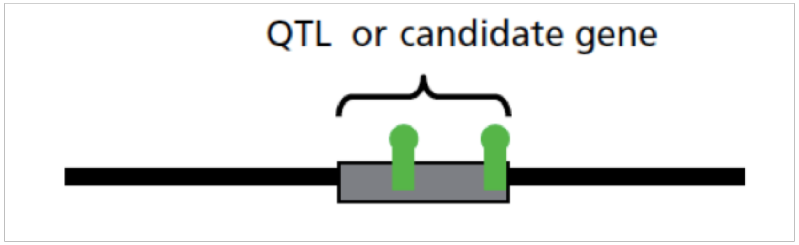
GEN E = QTL con efecto aditivo (el que quiero encontrar).

SNP C: Marcador en el genoma (el que tengo en mi CHIP).

EQUILIBRIO DE LIGAMIENTO
r MAX = 1/2

Genotipo en los ganetos (semen u ovas)	Frecuencia esperada	Freq. Esperada en 100 gametos
CE	½ (1-r)	25
ce	½ (1-r)	25
cE	½ r	25
Ce	½ r	25

DESEQUILIBRIO DE LIGAMIENTO



DESEQUILIBRIO DE LIGAMIENTO
Ej. r = 0,1

Genotipo en los ganetos (semen u ovas)	Frecuencia esperada	Freq. Esperada en 100 gametos
CE	½ (1-r)	45
ce	½ (1-r)	45
cE	½ r	5
Ce	½ r	5

QTL E - MARCADOR C - VALOR GENÉTICO a

Genotipo en los gametos	Freq. esperada 100 gametos	Valor genético	FREC x G
CE	45	½ +a	45 ½ a
ce	45	½ -a	- 45 ½ a
cE	5	½ +a	5 ½ a
Ce	5	½ -a	- 5 ½ a

C_{hijos}
40 ½ a ÷ 50

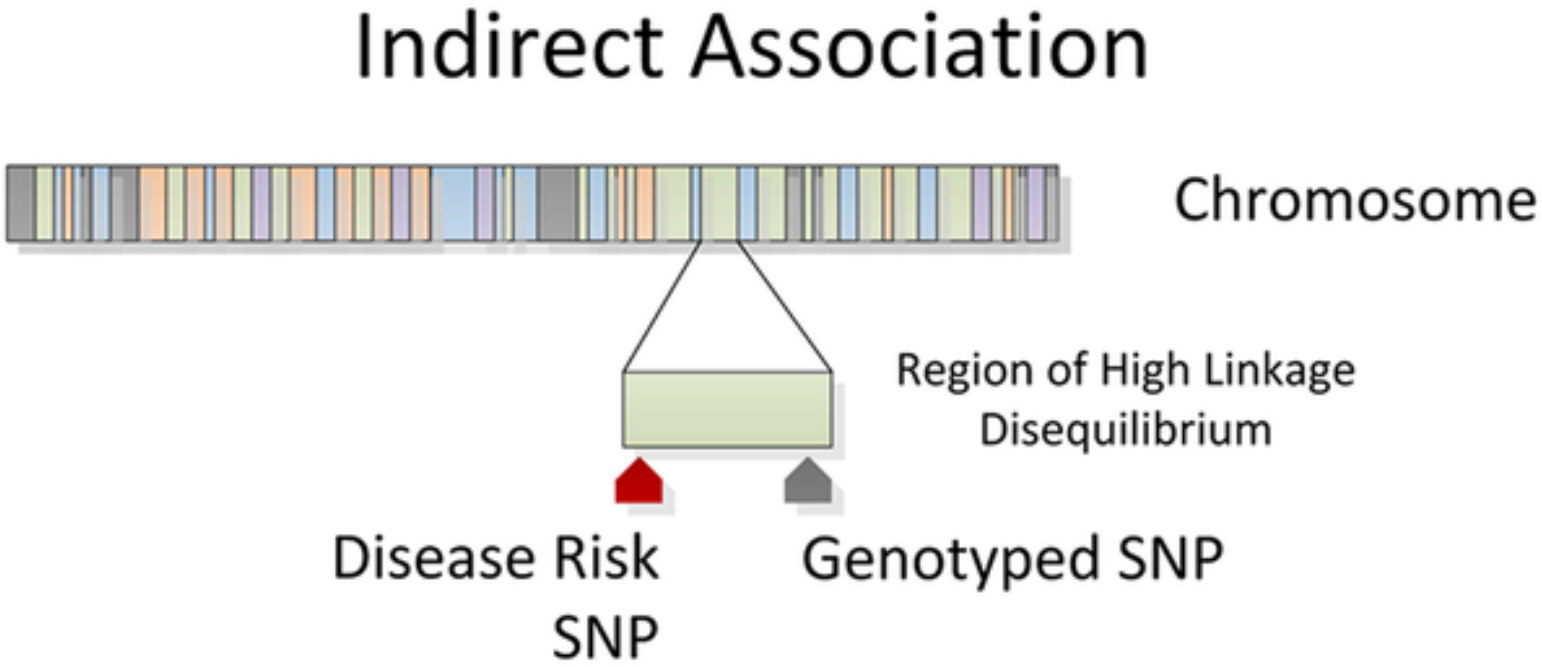
c_{hijos}
- 40 ½ a ÷ 50

Diferencia: $C_{hijos} - c_{hijos} = a * 40/50 = a*(1-2r)$

La diferencia depende de:
- El efecto del QTL = a
- La tasa de recombinación entre el marcador y el QTL = r.
- Si r = ½, el marcador no esta ligado al QTL y por lo tanto no hay diferencia entre los hijos que heredan el marcador C o c

Referencia: Theo Meuwissen (2010)

REPRESENTACIÓN DESEQUILIBRIO DE LIGAMIENTO EN UN CROMOSOMA (SNP-QTL)



Solo los marcadores altamente ligados al QTL podrán predecir el fenotipo.

Bush WS, Moore JH (2012)

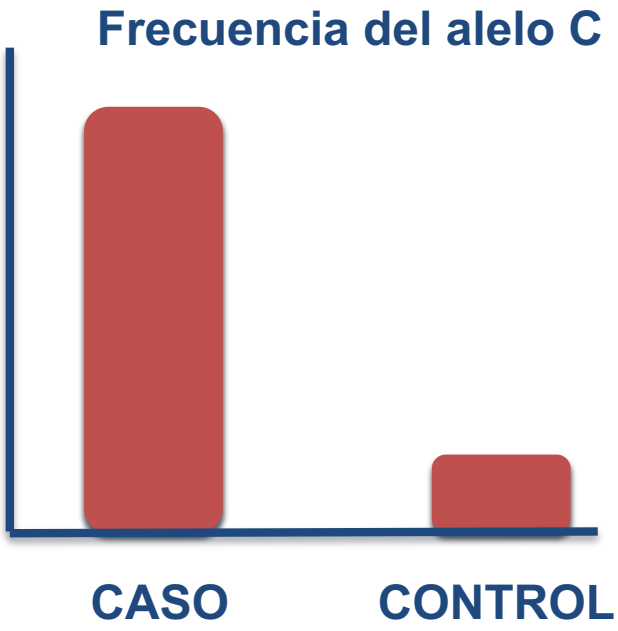
DISEÑO CASO - CONTROL

CASO

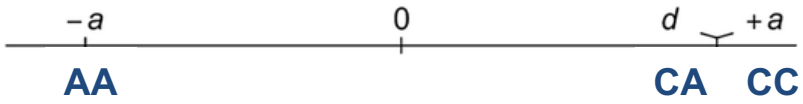
C	C

CONTROL

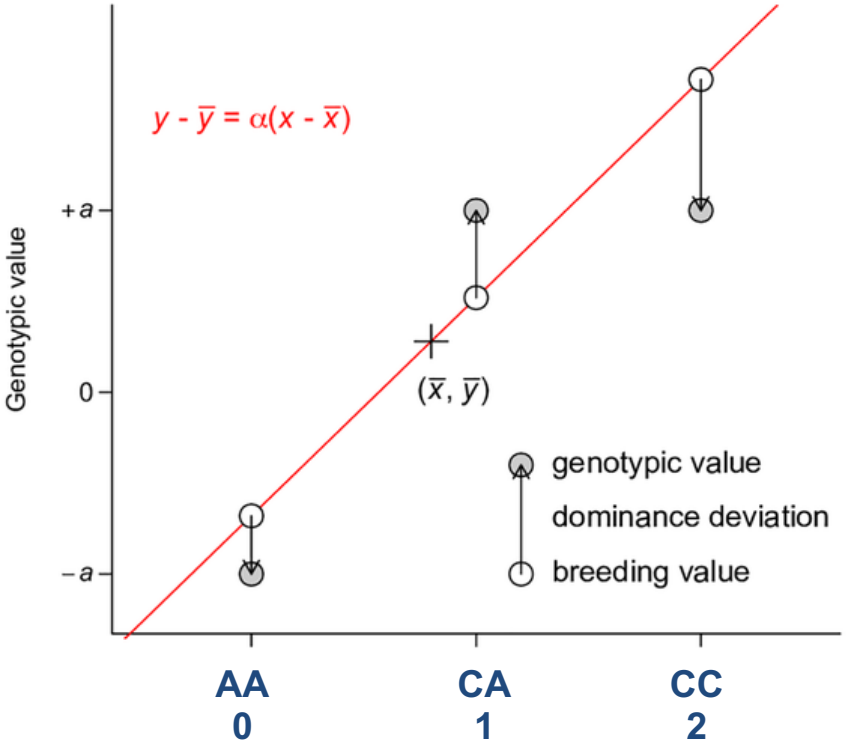
C	A
A	A



DISEÑO MODELO REGRESIÓN LINEAL (LOGÍSTICA) CON COVARIABLES



CC	CC	CC	CC	CA	CA	CA	AA	AA	AA
2	2	2	2	1	1	1	0	0	0



¿QUÉ DATOS NECESITO PARA UN ESTUDIO GWAS?

- **Una población con variación genética en <u>el rasgo de interes</u>.**

- **<u>Miles</u> de marcadores en <u>miles de individuos con registro de fenotipo</u>.**

Individual	M1	M2	M3	M4	Phenotype
1	2	2	2	1	20
2	1	0	0	0	8
3	2	2	0	0	24
4	1	1	2	1	4
5	1	0	0	0	14
6	1	1	2	1	12

PLOS Genetics 12(11): e1006421.

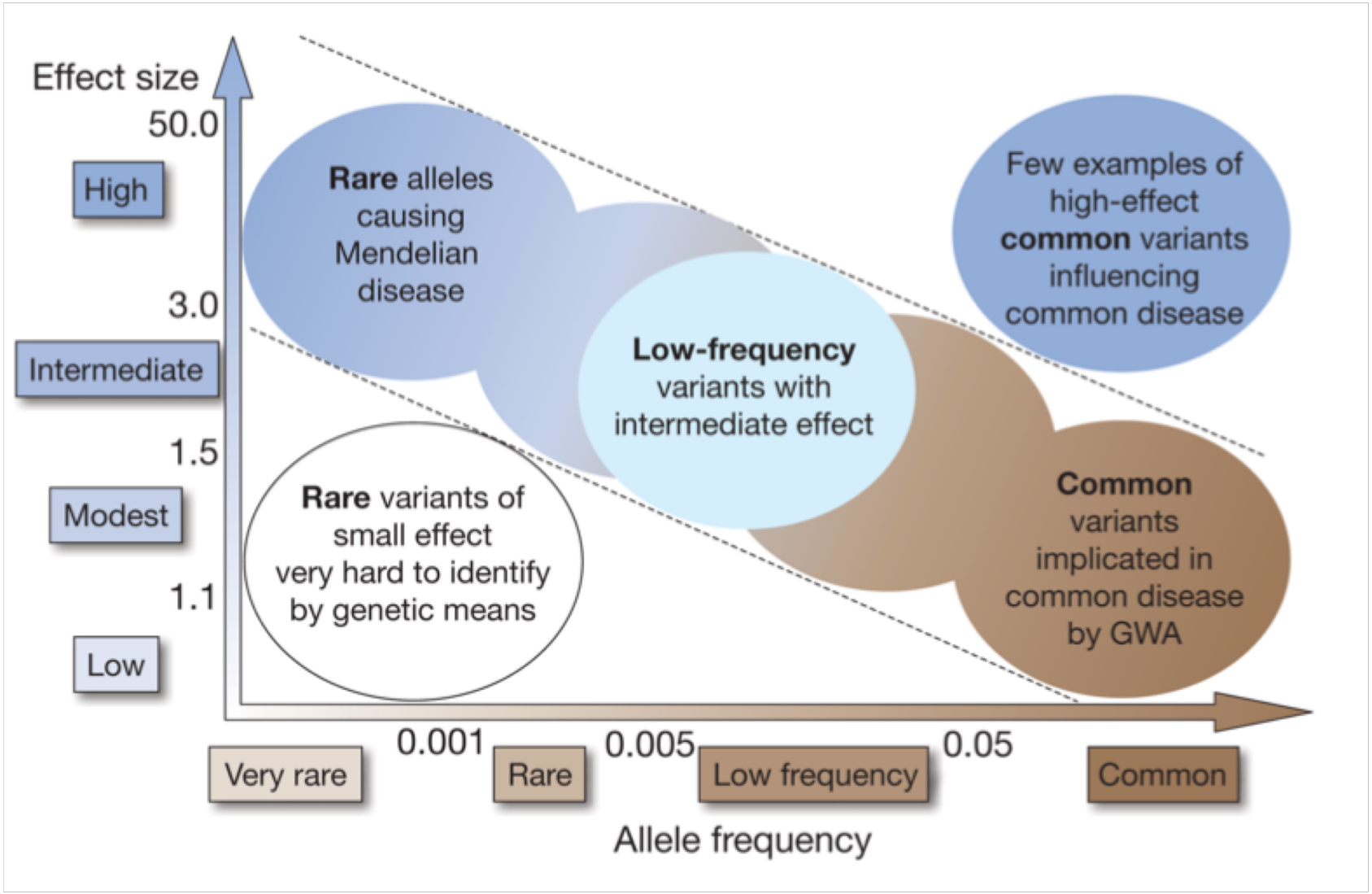
CONSIDERACIONE METODOLÓGICAS

1.- Buscar estructura poblacional: Si existe se debe incluir como covariable los Componentes Principales más importantes. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006).

2.- Filtrar parientes cercanos permite corregir posibles efectos de estructura familiar intrapoblacional.

3.- ¿Cuándo es significativo? Cuando su P value es menor a 5×10^{-8}, asumiendo que se realizan un millón de test de asociación independientes y se aplica un factor de Corrección de pruebas multiples (Bonferroni multiple test correction)

¿PUEDO ENCONTRAR TODOS LOS QTL?



Nature. 2009 October 8; 461(7265): 747–753.

SELECCIÓN ARTIFICIAL TRADICIONAL

Base datos Pedigree			Base datos producción	Base datos Breeding value
Sire	Dam	Animal	Phenotype	Breeding value
1	3	IP_9	20	-20
1	4	IP_10	30	-10
2	5	IP_11	40	+10
2	6	IP_12	50	+20
1	7	REPRO_13	?	-20
2	8	REPRO_14	?	+20
1	3	REPRO_15	?	0 *
1	3	REPRO_16	?	0 *

* Breeding value de hermanos de padre y madre es el mismo.

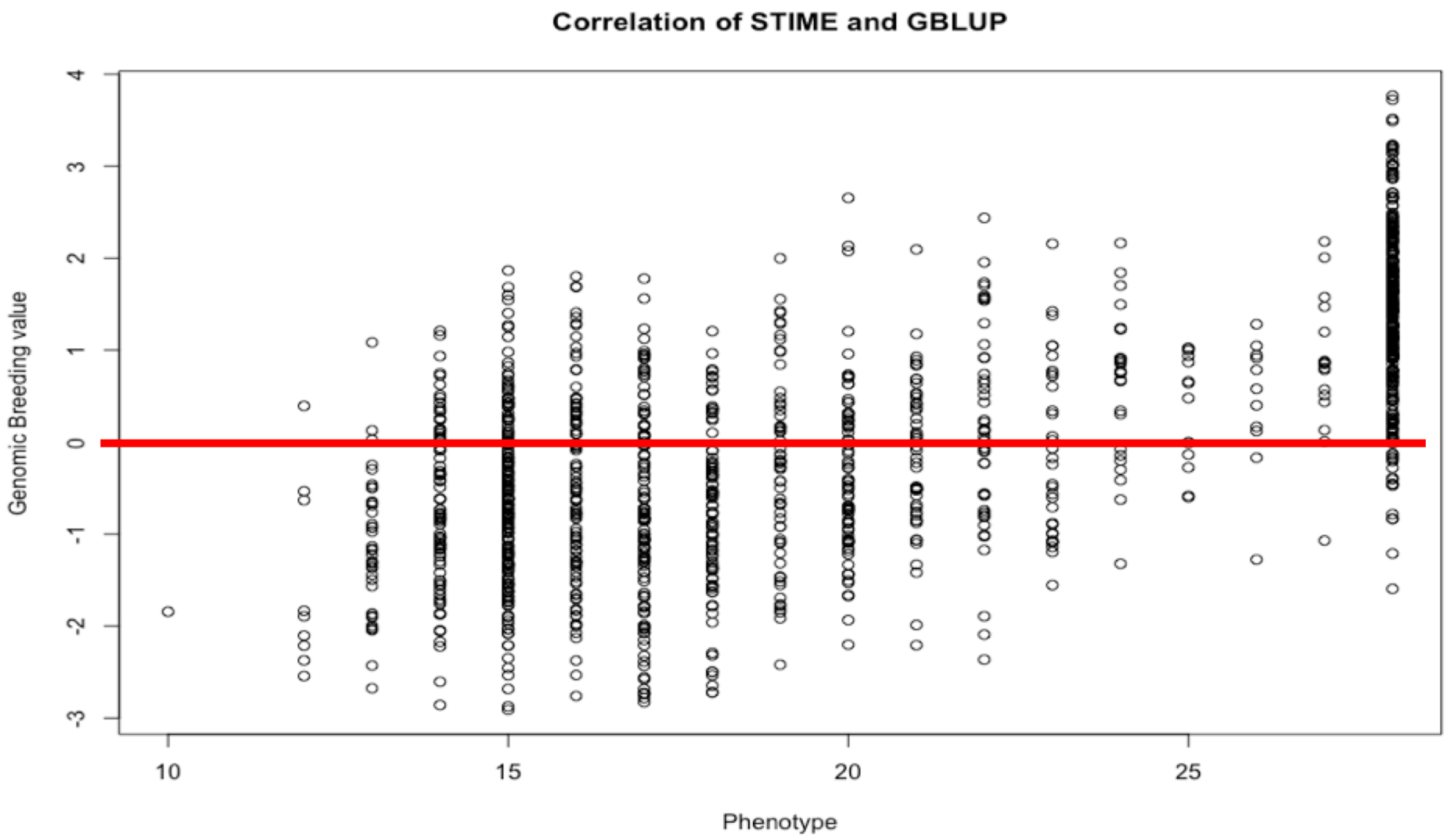
SELECCIÓN GENÓMICA: PREDICCIÓN DEL FENOTIPO SOLO CON EL GENOTIPO

Sire	Dam	Animal	SNP 1	SNP 2	SNP 3	SNP ….	Pheno	Genomic Breeding value
1	3	IP_9	0	0	2	0	20	-20
1	4	IP_10	0	1	1	0	30	-10
2	5	IP_11	1	1	2	0	40	+10
2	6	IP_12	0	0	1	0	50	+20
1	7	REP_13	0	2	2	0	?	-20
2	8	REP_14	0	1	2	0	?	+20
1	3	REP_15	1	2	1	0	?	-10 *
1	3	REP_16	0	1	2	0	?	+10 *

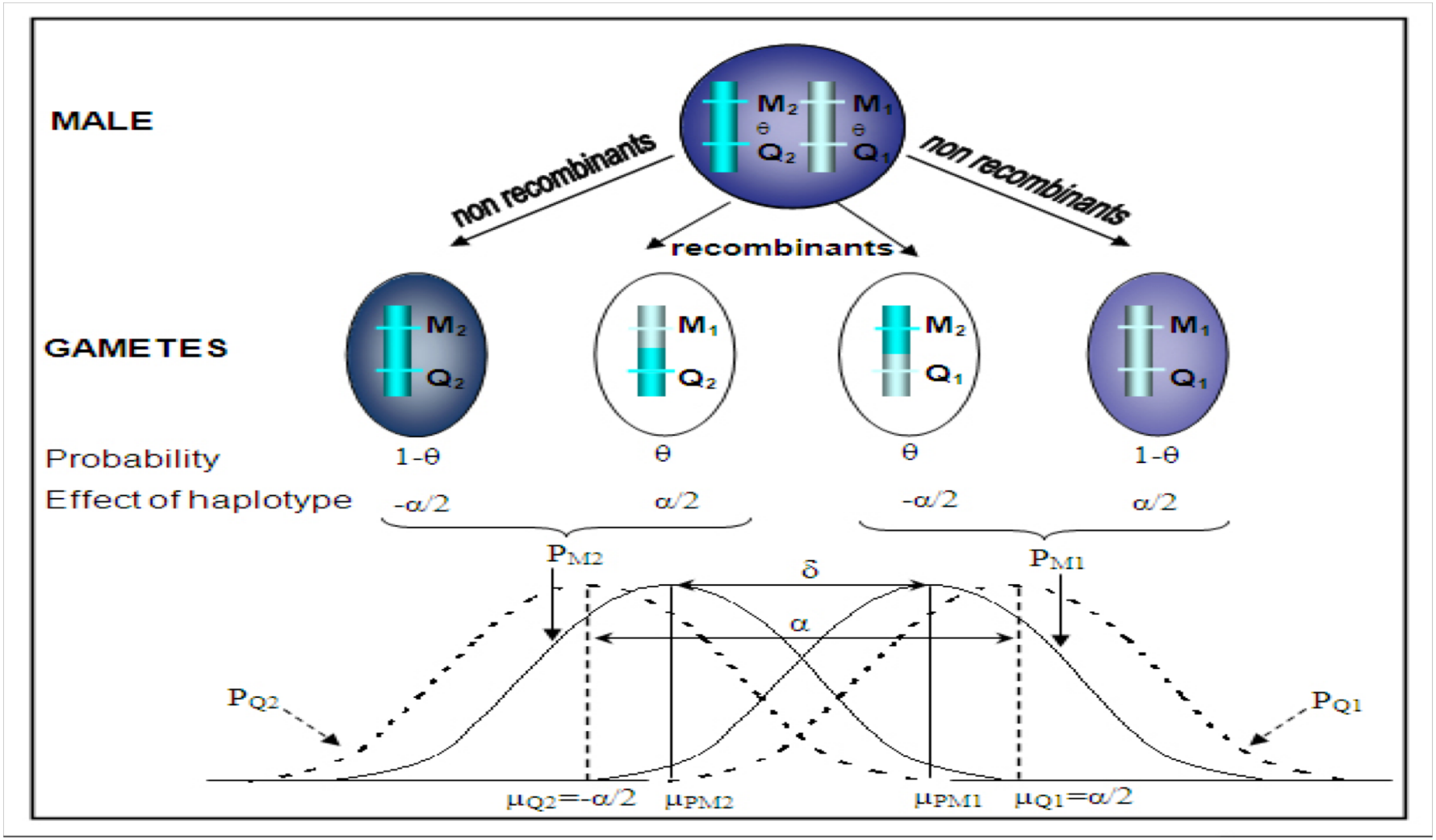
0 = 0 alelos
1 = 1 alelo
2 = 2 alelos

- Breeding value de hermanos de padre y madre es distinto permitiendo mejorar la predicción del breeding value.

EJEMPLO BREEDING VALUE GENÓMICO

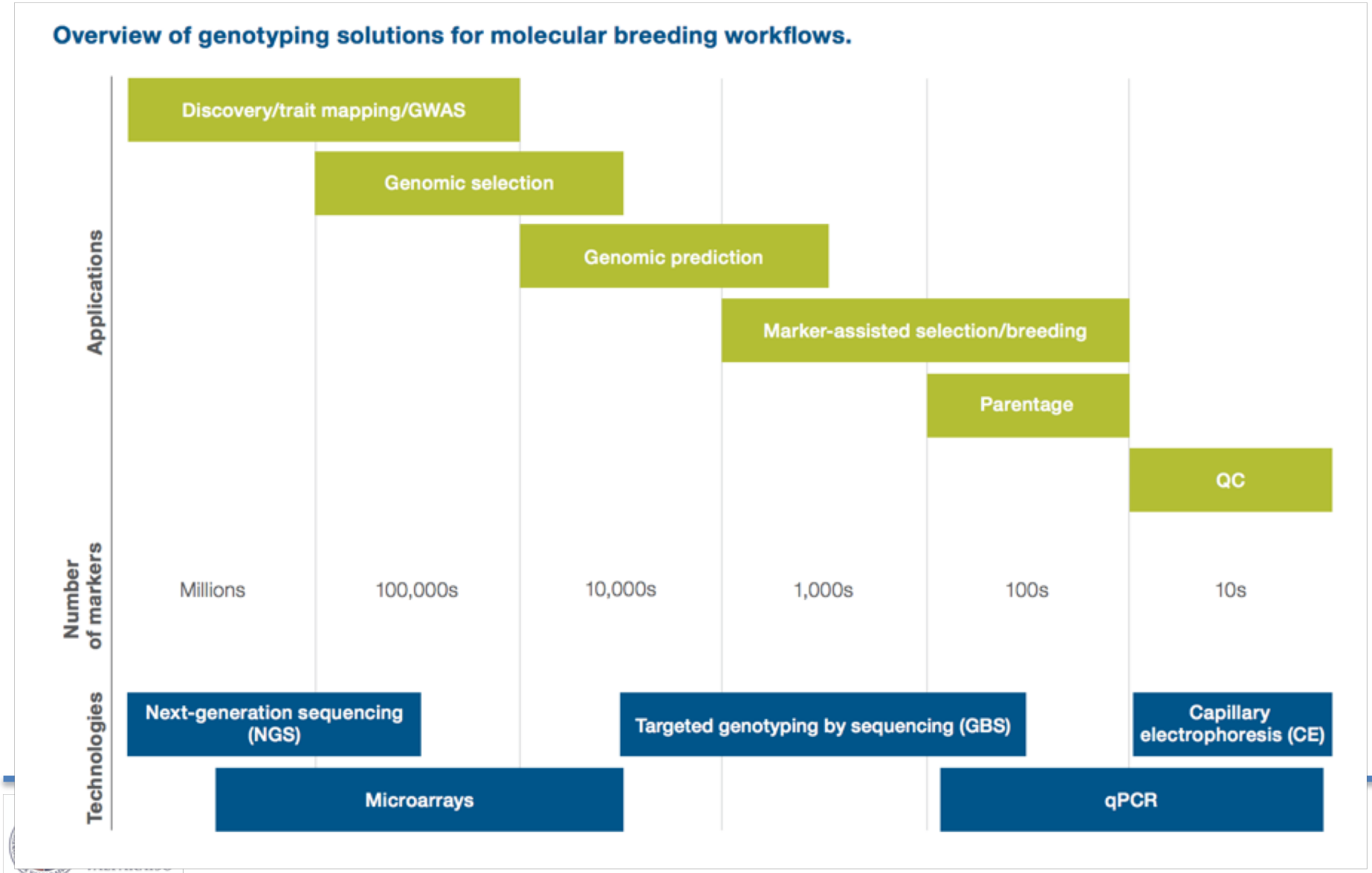


SELECCIÓN DE REPRODUCTORES QTL+/QTL- ¿CÓMO SON LO HIJOS?

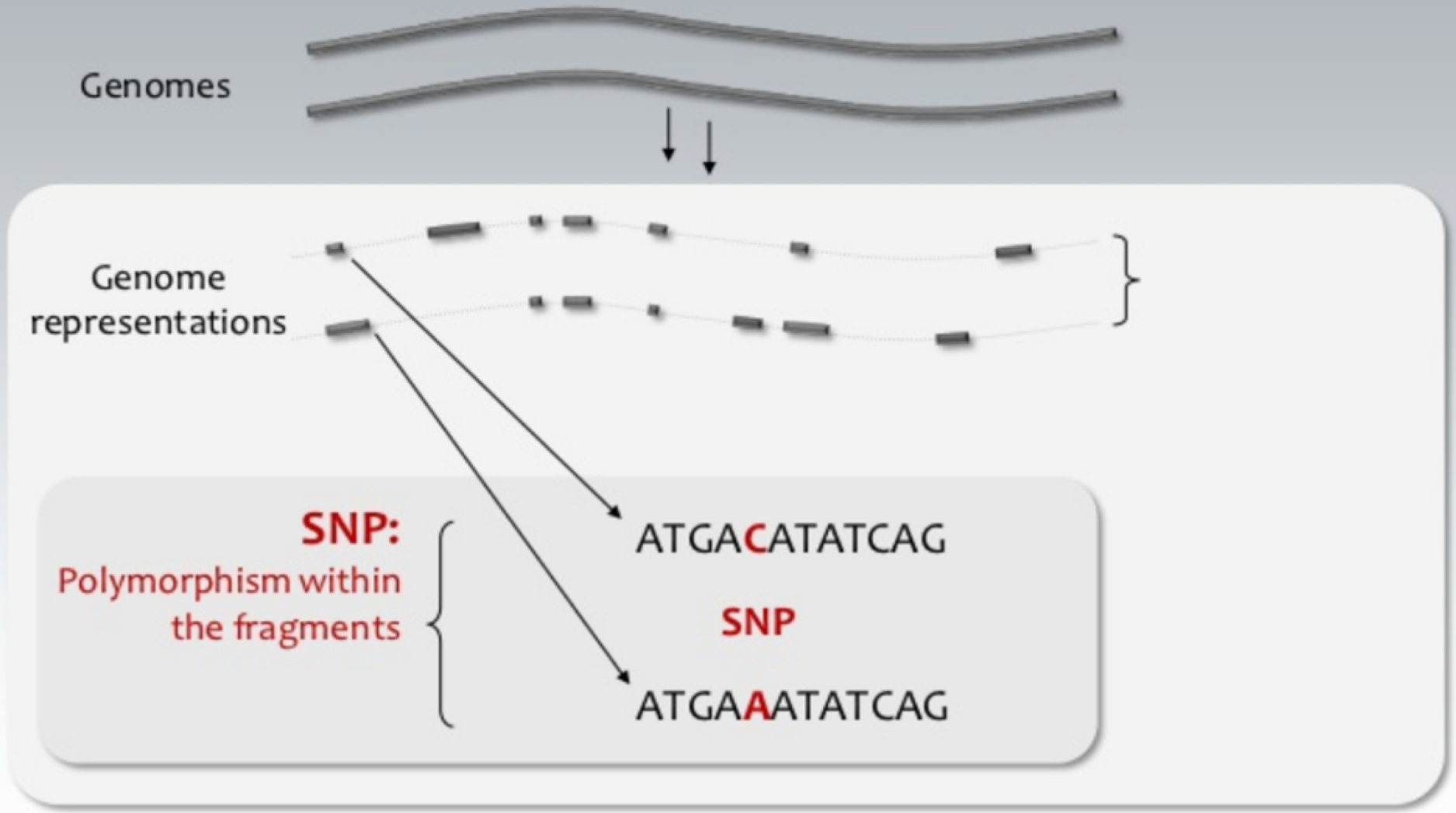


GENOTIPIFICACIÓN

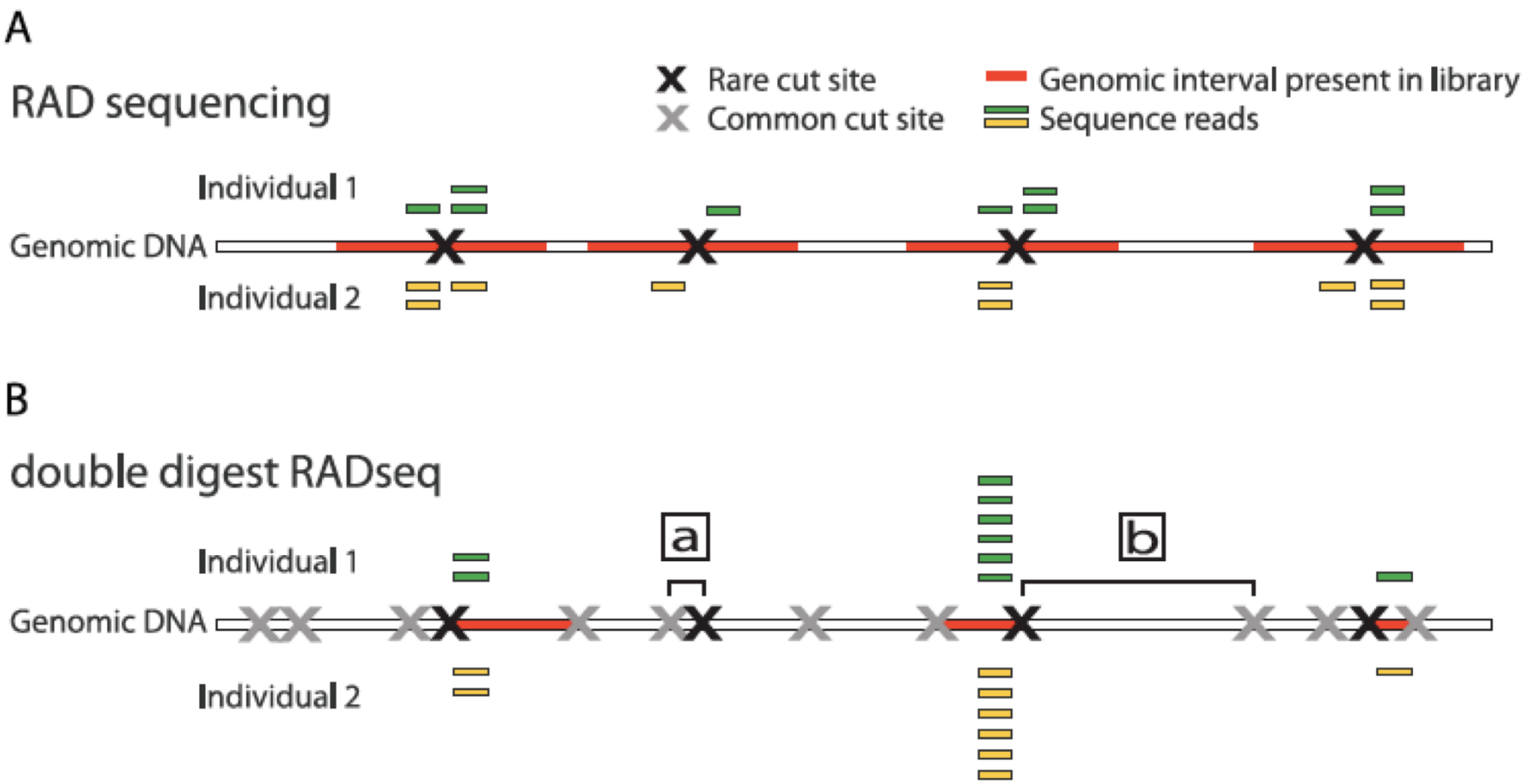
TÉCNICAS DE GENOTIPIFICACIÓN EN FUNCIÓN DEL TAMAÑO



Genotyping-by-sequencing (GBS)



GENOTIPIFICACIÓN MASIVA *DE NOVO*
*RAD-seq (*Restriction-site-associated DNA sequencing)



GENOTIPACIÓN MASIVA:
Axiom™ Salmon Genotyping Array.

ThermoFisher SCIENTIFIC

- >130,000 SNPs evenly spaced across the Atlantic salmon genome.
- 87 Y chromosome-specific probes for accurate gender determination.

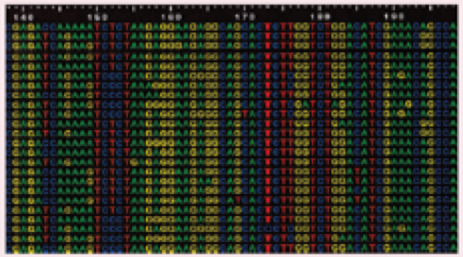
BD.SNP

ANIMAL	SNP 1	SNP 2	SNP 3	SNP ….	SNP 130.000
IP_9	GG	TT	CC	AA	AG
IP_10	GG	TC	TC	AA	AG
IP_11	GG	TC	CC	AA	GG
IP_12	GG	TT	TC	AA	GG
REP_13	GG	CC	CC	AA	GG
REP_14	GG	TC	CC	AA	GG

GENOTIPACION MASIVA CON CHIPs

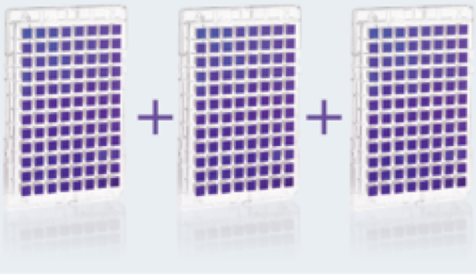
SNP discovery
Whole-genome sequencing

- Select SNPs for screen based on:
  - Likelihood to be polymorphic
  - MAF
  - Sequencing quality scores
  - Presence in multiple populations
  - Associations to traits of interest (if known)



SNP validation
1.3M–2M SNP screening

- Select SNPs for routine testing based on:
  - High performance in assay
  - Polymorphic in multiple populations
  - Genome spacing: LD, imputation, or physical density
  - Higher marker density in genes of interest



High-volume genotyping
1.5K–650K SNPs

- Genotyping with most informative markers
- Highly powered with thousands of samples
- Well suited for routine testing, marker-trait associations, QTL mapping

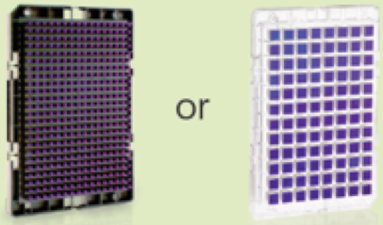
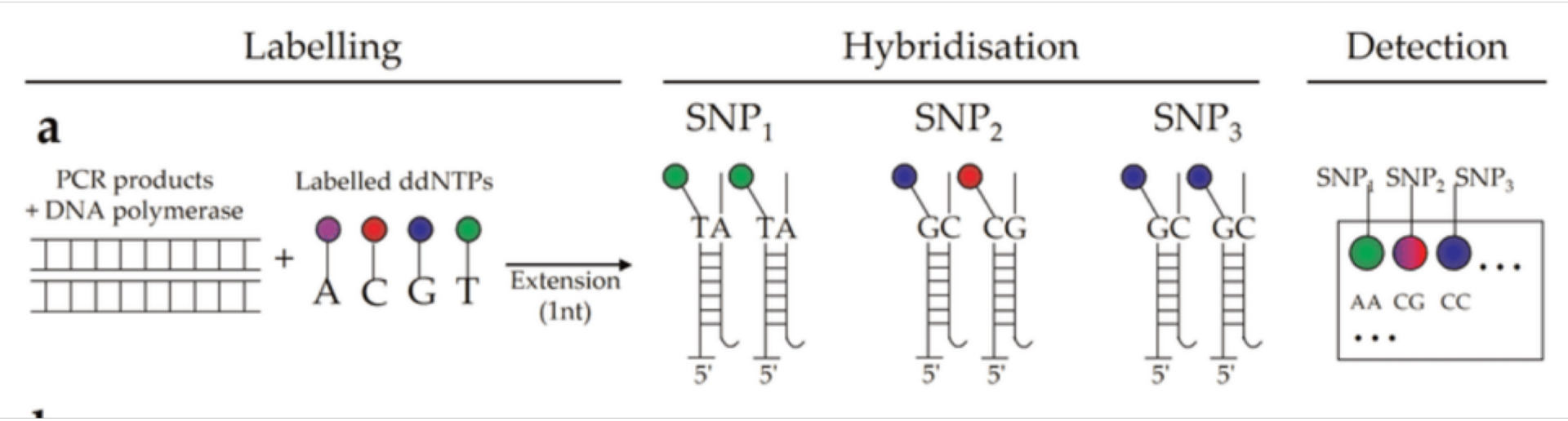


Figure 1. Process for SNP discovery with DNA sequencing, SNP validation with a large screening experiment on multiple arrays, and high-volume genotyping for downstream discoveries and routine testing.

GENOTIPACION MASIVA CON CHIPs

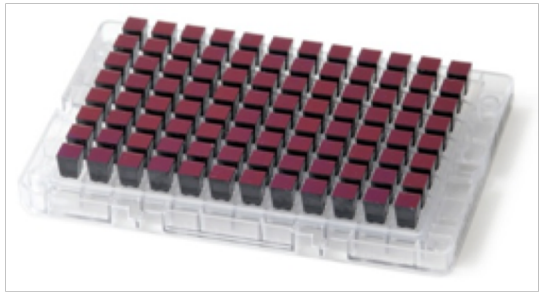


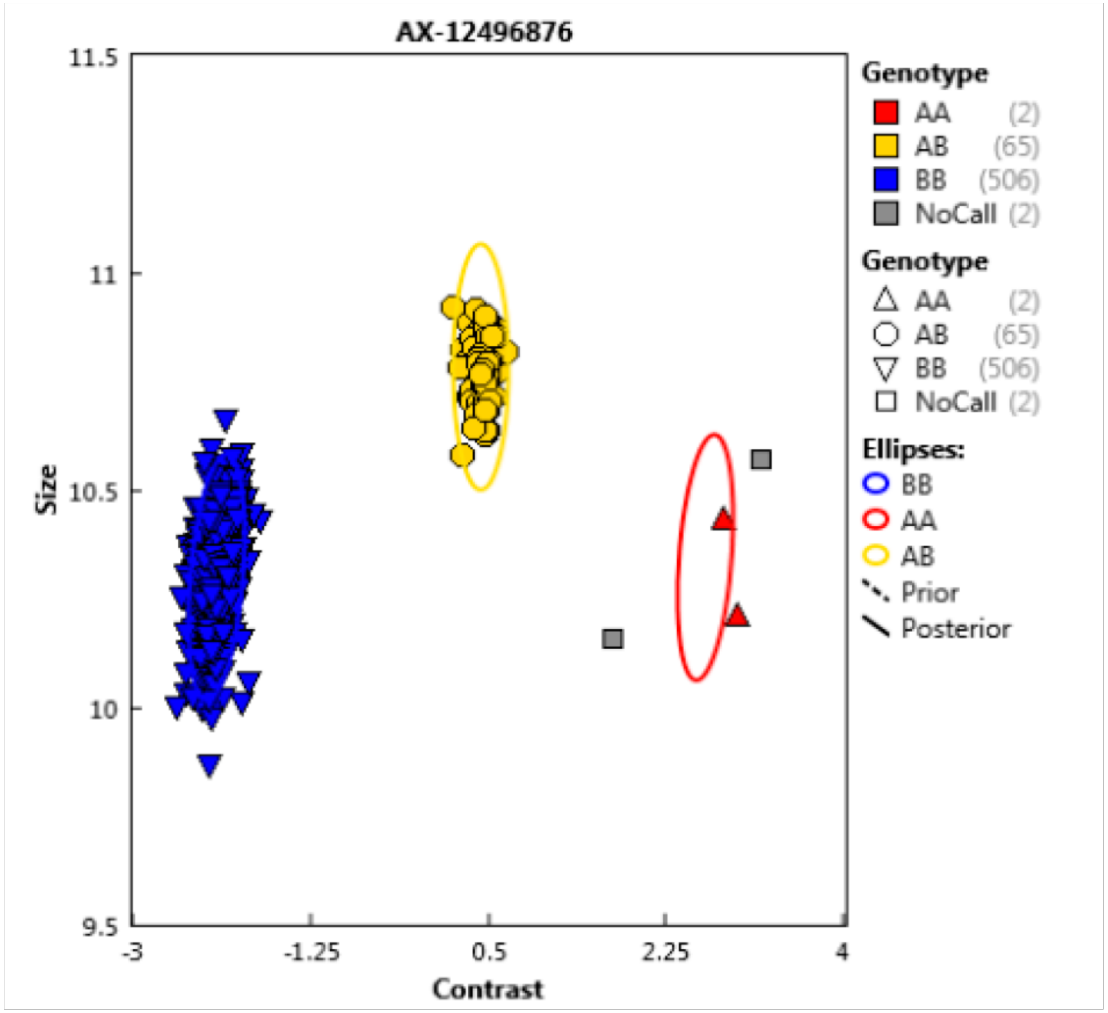
A streamlined assay

Total genomic DNA (200 ng) is amplified and randomly fragmented into 25 to 125 base pair (bp) fragments. These fragments are purified, re-suspended, and hybridized to Axiom Genome-Wide Human Array Plates. Following hybridizatio the bound target is washed under stringent conditions to remove non-specific background to minimize background noise caused by random ligation events. Each polymorphic nucleotide is queried via a multi-color ligation event carried out on the array surface. After ligation, the arrays are stained and imaged on the GeneTitan™ Multi-Channel Instrument.

RESULTADO ESPERADO

ThermoFisher SCIENTIFIC





RESULTADO OBSERVADO Y FILTRO DE CALIDAD

Figure 14. Examples of a well-clustered SNP (on the left) and mis-clustered SNP (on the right).

Examples of well-clustered SNP and mis-clustered SNPs, SNP Metrics that detect mis-clustering are red

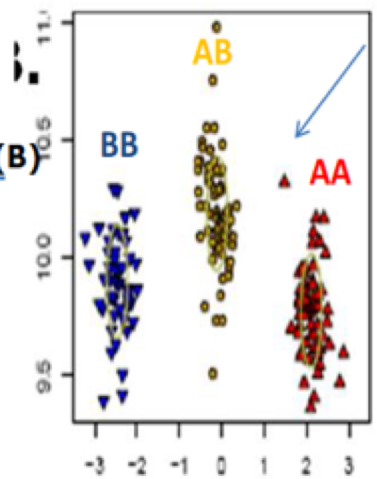
Well-clustered

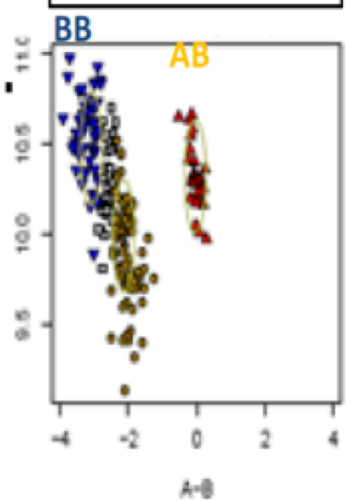
Each pt is a sample

Mis-clustered

Size:

$$\frac{\text{Log}_2(A)+\text{Log}_2(B)}{2}$$





Contrast:
$\text{Log}_2(A/B)$

SNP Call Rate,
SNP FLD,
HetSO

Will detect this

Called Genotypes:
AA(blue), AB(gold), BB (red)
NoCalls (grey)

EJEMPLO CHIP SALMON DEL ATLÁNTICO

affymetrix
Biology for a better world

Table 1: Techniques of DNA and RNA sequencing experiments used for SNP discovery

	RR-Seq	RAD-Seq	RNA-Seq
Samples	*Farmed (40), wild (16), haploid (1)	*Farmed (160)	*Farmed (72)
Initial discovered SNPs	472,072	467,268	816,570
SNPs selected for array**	99,097	83,151	229,754
Final SNPs for array design***	73,800	54,197	156,979

*Farmed fish were representative samples of the Atlantic salmon breeding company Landcatch Natural Selection Ltd.
**SNPs selected after filtering for PSVs, MAF, and repeat elements
***SNPs selected after calculation of *in silico* design scores

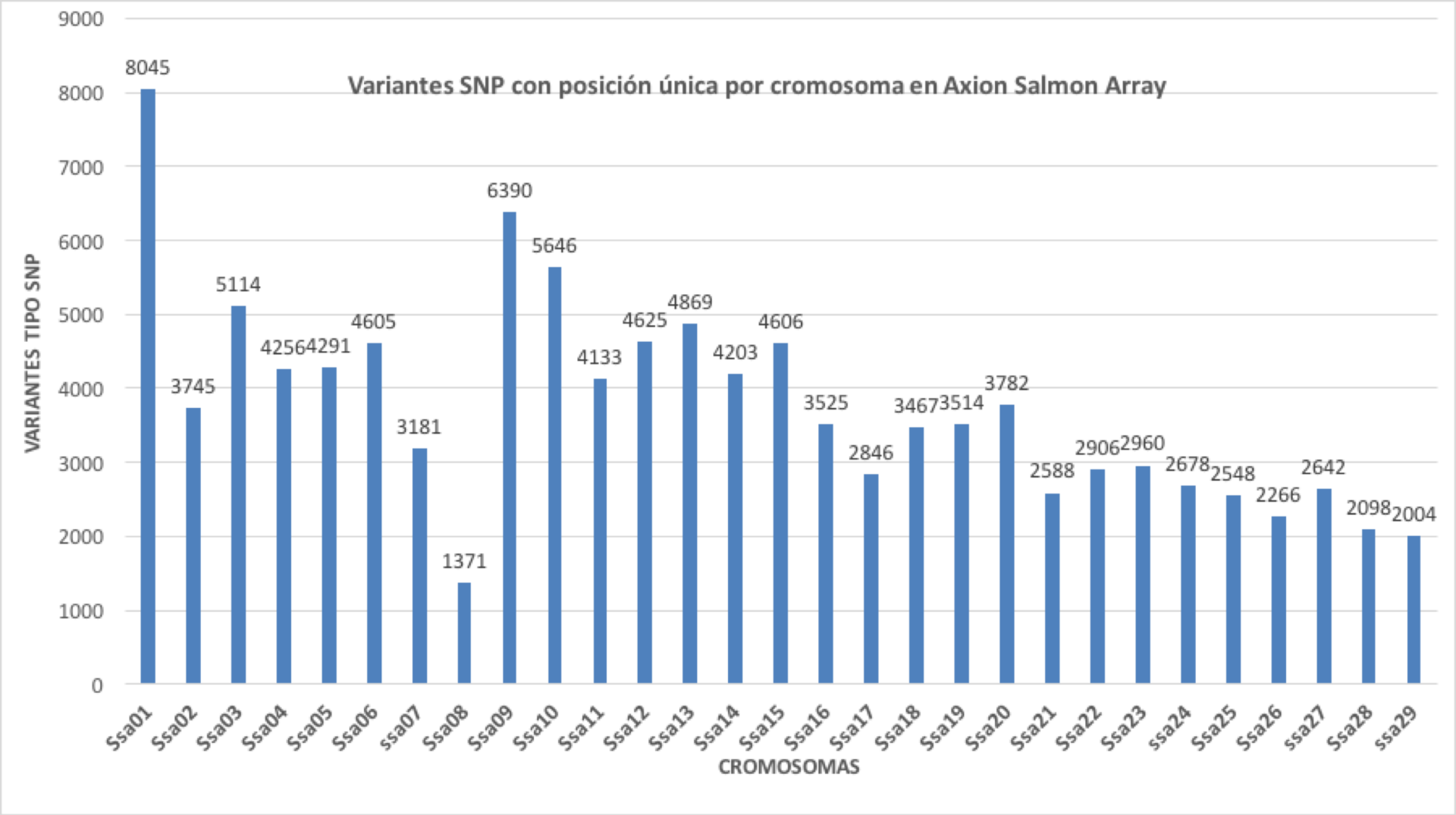
EJEMPLO MARCADORES SNP EN AXIOM SALMON ARRAY.

SNP_ID	Chromosome	Position	Type_variant	Genomic_Variant	Alternative_Variant
AX-87956229	Ssa01	4545	3_prime_UTR	C	G
AX-88260982	Ssa01	6115	3_prime_UTR	T	G
AX-88135017	ssa01	228680	3_prime_UTR	G	T
AX-88217386	ssa01	1237122	3_prime_UTR	C	T
AX-88087518	ssa01	2280167	3_prime_UTR	C	T
AX-87878666	ssa01	2940121	3_prime_UTR	G	A
AX-87939100	ssa01	4007497	3_prime_UTR	A	C
AX-88277162	ssa01	4303515	3_prime_UTR	G	A
AX-88027426	ssa01	4436890	3_prime_UTR	T	C
AX-88109535	ssa01	4474162	3_prime_UTR	A	T
AX-88280057	ssa01	4524218	3_prime_UTR	C	T
AX-88127212	ssa01	4549000	3_prime_UTR	C	G
AX-88108785	ssa01	4549095	3_prime_UTR	G	A

ANOTACIÓN FUNCIONAL
SNP - AXIOM SALMON ARRAY.

Tipo de función	Nº variantes posición única
intergenic_region	58936
intron	34879
3_prime_UTR	6331
synonymous	3407
missense	2527
5_prime_UTR	1889
multiple	519
start_gain	330
splice_region	34
stop_gained	33
stop_lost	13
stop_retained	5
start_lost	1
Total	**108904**

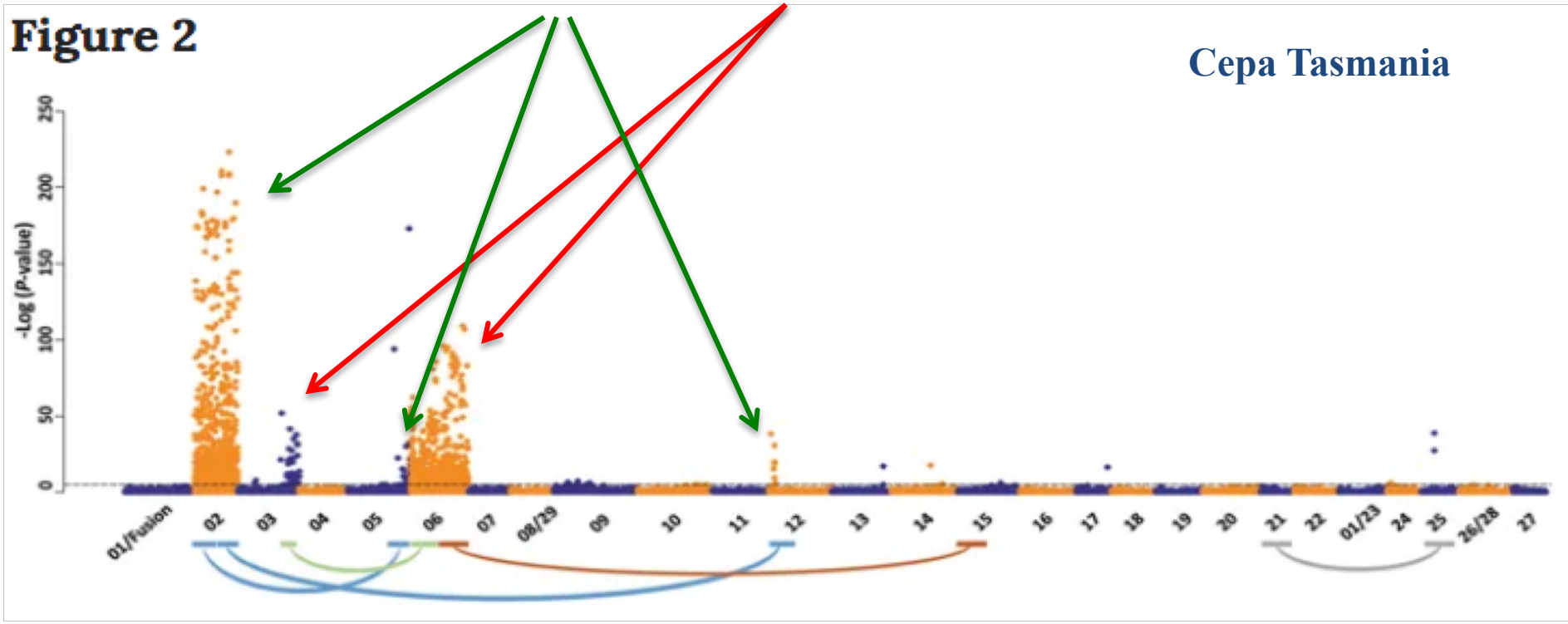
VARIENTES POR CROMOSOMA
SNP - AXIOM SALMON ARRAY.



ESTUDIOS DE CASO GWAS

Estudio determinación cromosoma sexual Salmón del Atlántico (Kijas et al 2018).

2 linajes de localización del gen determinante del sexo Crom 2/5/12 y Crom 3/6.



ESTUDIO DE CASO QTL CRECIMIENTO

BMC Genet. 2013; 14: 112. PMCID: PMC4219356
Published online 2013 Nov 27. doi: 10.1186/1471-2156-14-112 PMID: 24283985

A SNP in the 5′ flanking region of the myostatin-1b gene is associated with harvest traits in Atlantic salmon (Salmo salar)

Carolina Peñaloza,[1] Alastair Hamilton,[2] Derrick R Guy,[2] Stephen C Bishop,[1] and Ross D Houston [1]

Table 2

Genotype and allele frequencies of SNPs

	g.1060C>A			*g.1086C>T*			*c.3501C>G*		
Genotype	AA	AC	CC	TT	TC	CC	CC	CG	GG
Frequency	74	1128	3474	361	1951	2348	3186	1392	102
Minor allele frequency		0.14			0.29			0.17	
Major allele frequency		0.86			0.71			0.83	

Asociado significativamente con rasgos de peso corporal.

https://europepmc.org/article/med/24283985

ESTUDIO DE CASO QTL CRECIMIENTO

Table 4

Genotype means with standard error, additive effect and dominance effects for the SNP that showed a significant trait association

	SNP	Genotype Mean ± SE [1]				
Trait	***g.1086C > T***	**TT**	**TC**	**CC**	***a* ± SE**	***d* ± SE**
HWT		2.60 ± 0.04	2.55 ± 0.02	2.50 ± 0.02	0.05 ± 0.017*	−0.0002 ± 0.02
GWT		2.37 ± 0.03	2.33 ± 0.02	2.29 ± 0.02	0.04 ± 0.019*	0.003 ± 0.02
DHWT		2.07 ± 0.03	2.04 ± 0.01	2.00 ± 0.01	0.03 ± 0.017*	0.007 ± 0.01
FLWT		1.73 ± 0.02	1.69 ± 0.01	1.66 ± 0.01	0.03 ± 0.014*	−0.007 ± 0.01

[1]Estimate of the effect is expressed in kg.

*$p < 0.05$.

SNP	Secuencia
g.1086C>T	TCCTTGTATC AAAC[C/T]TTTCT AAGGCTACCA

ESTUDIO DE CASO
QTL *P. salmonis*

BMC Genomics. 2015; 16: 854.
Published online 2015 Oct 24. doi: 10.1186/s12864-015-2038-7
PMCID: PMC4619534
PMID: 26499328

Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes

Katharina Correa, Jean P. Lhorente, María E. López, Liane Bassini, Sudhir Naswa, Nader Deeb, Alex Di Genova, Alejandro Maass, William S. Davidson, and José M. Yáñez

Table 2
Minor allele frequency of each marker and effect of each allele on the resistance against *Piscirickettsia salmonis*

SNP id	A_1	A_2	MAF	$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	Effect (A_2)
AQI_UCh-93346943	C	A	0.24	34.96	32.77	31.33	−2.15 (0.26)
AQI_UCh-93451692	C	A	0.29	34.64	32.27	29.34	−2.91 (0.32)
AQI_UCh-93323028	G	T	0.31	34.53	32.04	30.24	−2.73 (0.33)
AQI_UCh-93268476	C	T	0.24	35.40	33.50	31.86	−1.89 (0.23)
AQI_UCh-93384759	G	T	0.31	34.70	32.74	29.93	−2.31 (0.28)

A_1 and A_2 are the minor and major alleles, respectively. Mean value of the time to death by genotype for each marker. A_1: high resistance allele. Effect of the A_2 allele and standard error in brackets

OBJETIVOS DEL TRABAJO PRÁCTICO

Esta práctica tiene como propósito realizar un análisis de GWAS