



Ferramentas genômicas no melhoramento genético de tilápia e camarão

Yuri Tani Utsunomiya

Faculdade de Medicina Veterinária de Araçatuba
Departamento de Apoio, Produção e Saúde Animal



AgroPartners
CONSULTING





Araçatuba

Capital do boi gordo

Faculdade de Medicina Veterinária

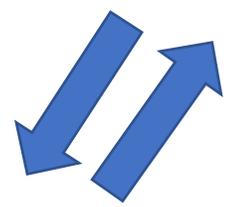
330 alunos de graduação
200 alunos de pós-graduação
50 docentes



unesp 

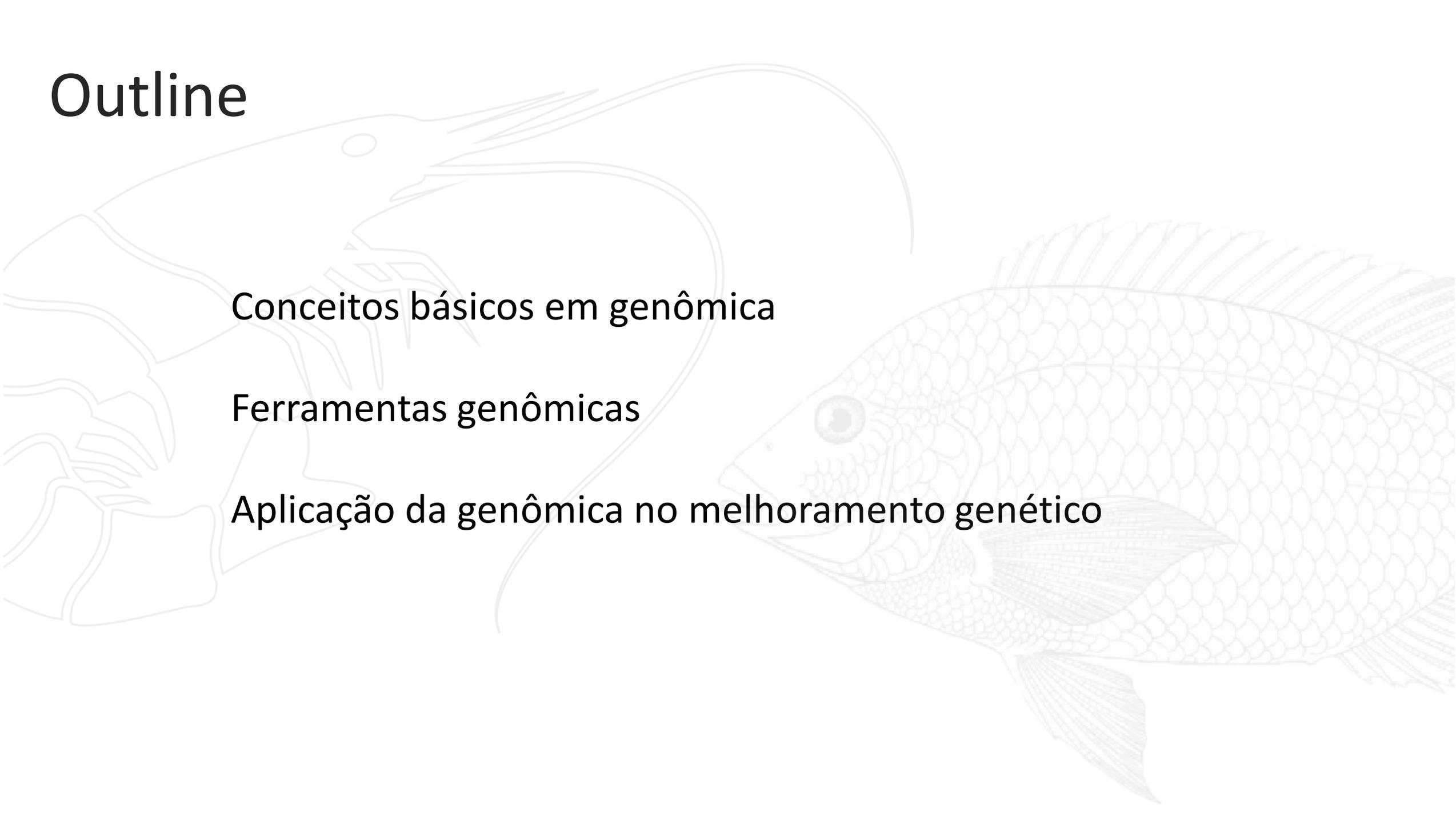


AgroPartners
CONSULTING



trisene 

Outline

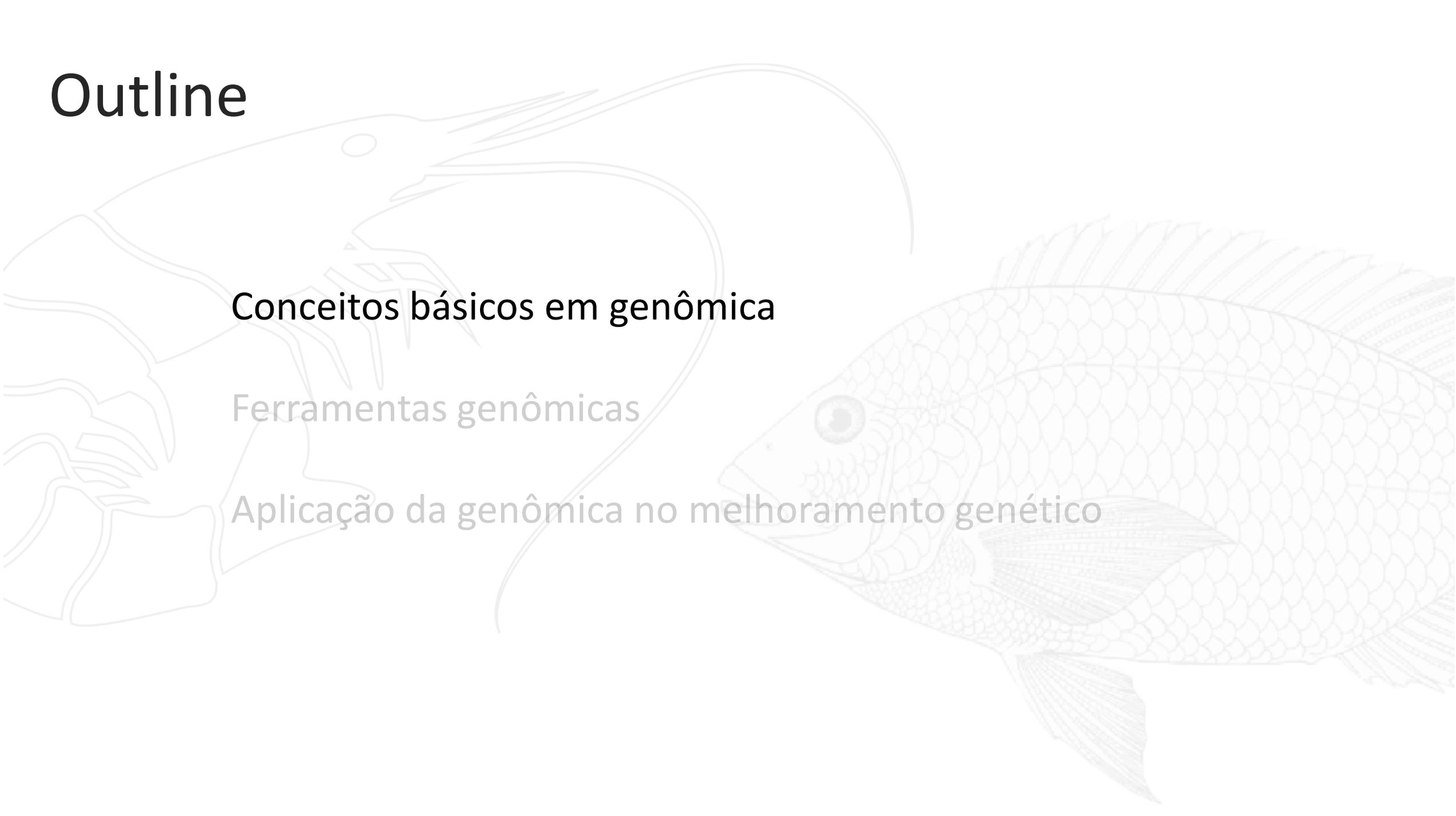


Conceitos básicos em genômica

Ferramentas genômicas

Aplicação da genômica no melhoramento genético

Outline



Conceitos básicos em genômica

Ferramentas genômicas

Aplicação da genômica no melhoramento genético

Variação fenotípica

$$F = G + A + (G \times A)$$

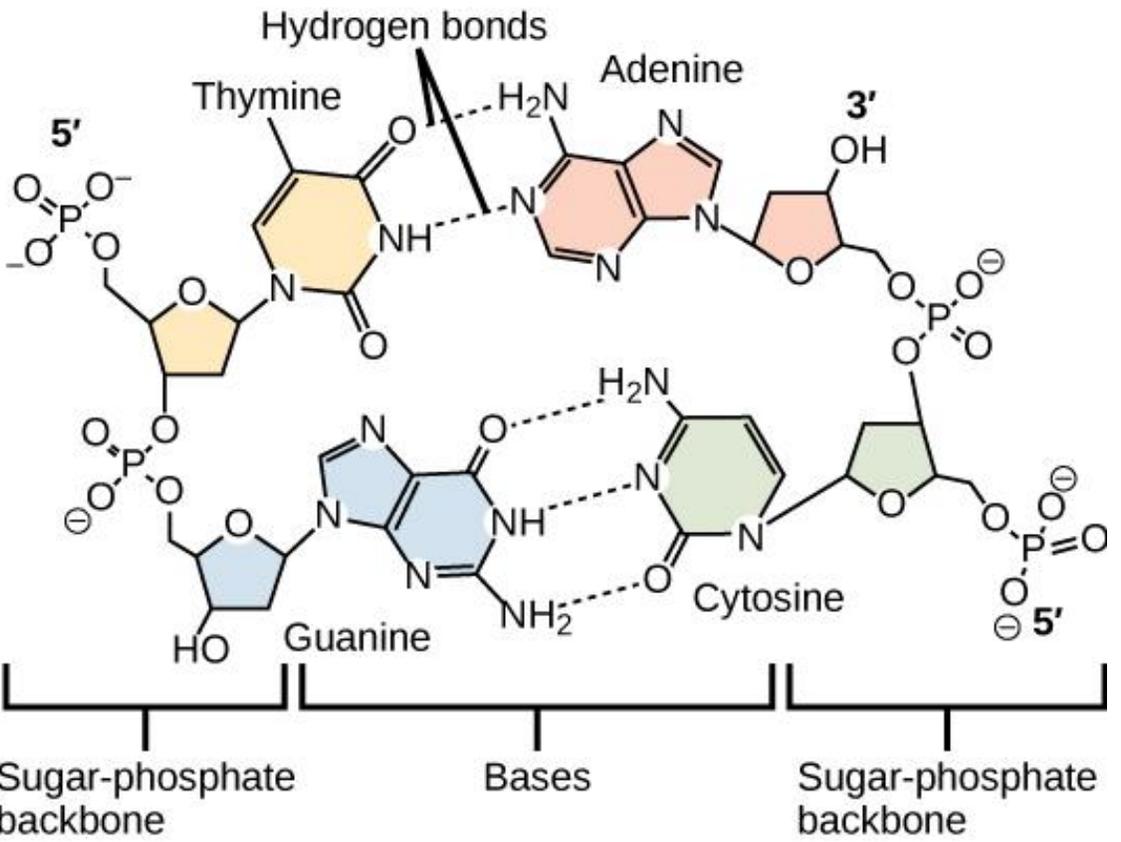
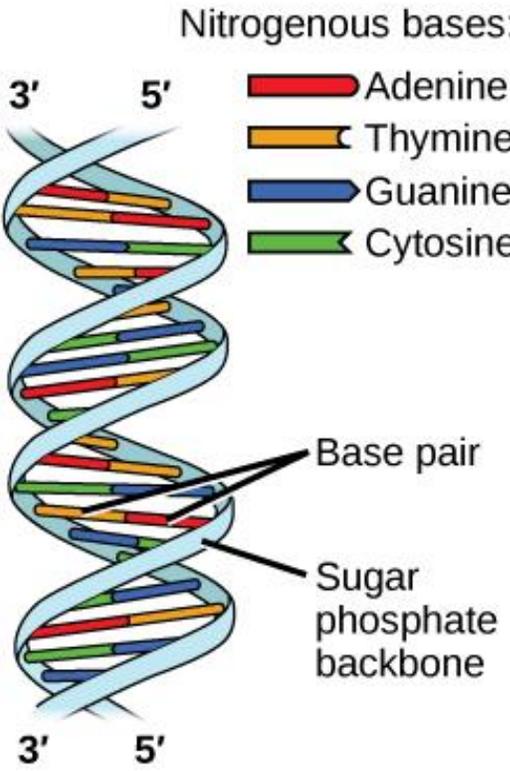
Fenótipo

Genoma
(DNA)

Ambiente

The diagram illustrates the equation for phenotypic variation: $F = G + A + (G \times A)$. The variable F is defined as the Phenotype (Fenótipo). The variable G is defined as the Genome (DNA) (Genoma (DNA)). The variable A is defined as the Environment (Ambiente). The interaction term $(G \times A)$ represents the combined effect of the genome and the environment.

Estrutura do DNA



Variantes genéticas

ATTCCGGATTAGTACGATCACCAATCGGTTCCGATGATATAGCATGCTA	Referência
ATTCCGG G TTAGTACGATCACCAATCGGTT C A A GATGATATAGCATGCTA	SNP
ATTCCGGGTTAGTACG GAG ACCAATCGGTT CAG ATGATATAGCATGCTA	MNP
ATTCCGGGTTAGTACGATCAC CCC AATCGGTTCCGATGA ----- ATGCTA	INDEL
ATTCCGG [ATT]₁₀ AGTACGATCACCAATCGGTTCCGATGATATAGCATGC	Microssatélite



SNP

Variante mais abundante

Variantes formam blocos (haplótipos)

ATTCCGG[ATT]₁₀AGTACGATCACCAATCGGTTCCGATGATATAGCATGC
ATTCCGG[ATT]₁₀AGTACGATCACCAATCGGTTCCGATGATATAGCATGC
ATTTCCGG[ATT]₁₅AGTACGATCACCGATCGGTTCCAATGATATACCATGC
ATTTCCGG[ATT]₁₅AGTACGATCACCGATCGGTTCCAATGATATACCATGC
ATTTCCGG[ATT]₁₂AGTACGATCACCAATCGGTTCCAATGATATAGCATGC
ATTCCGG[ATT]₁₅AGTACGATCACCAATCGGTTCCAATGATATAGCATGC
ATTCCGG[ATT]₁₅AGTACGATCACCAATCGGTTCCAATGATATAGCATGC
ATTCCGG[ATT]₁₅AGTACGATCACCAATCGGTTCCAATGATATAGCATGC

Variantes formam blocos (haplótipos)

ATTCCGG[ATT]₁₀AGTACGATCACCAATCGGGTTCCGATGATATAGCATGC
ATTCCGG[ATT]₁₀AGTACGATCACCAATCGGGTTCCGATGATATAGCATGC
ATTCGG[ATT]₁₅AGTACGATCACCGATCGGGTTCCAATGATATACCATGC
ATTCGG[ATT]₁₅AGTACGATCACCGATCGGGTTCCAATGATATACCATGC
ATTCGG[ATT]₁₂AGTACGATCACCAATCGGGTTCCAATGATATAGCATGC
ATTCGG[ATT]₁₅AGTACGATCACCAATCGGGTTCCAATGATATAGCATGC
ATTCGG[ATT]₁₅AGTACGATCACCAATCGGGTTCCAATGATATAGCATGC
ATTCGG[ATT]₁₅AGTACGATCACCAATCGGGTTCCAATGATATAGCATGC

Variantes formam blocos (haplótipos)

C[ATT]₁₀AGG

Haplótipo 1 (25%)

C[ATT]₁₀AGG

T[ATT]₁₅GAC

Haplótipo 2 (25%)

T[ATT]₁₅GAC

T[ATT]₁₂AAG

Haplótipo 3 (12,5%)

C[ATT]₁₅AAG

C[ATT]₁₅AAG

Haplótipo 4 (37,5%)

C[ATT]₁₅AAG

Variantes formam blocos (haplótipos)

C[ATT]₁₀AGG Haplótipo 1 (25%)

T[ATT]₁₅GAC Haplótipo 2 (25%)

T[ATT]₁₂AAG Haplótipo 3 (12,5%)

C[ATT]₁₅AAG Haplótipo 4 (37,5%)

Variantes formam blocos (haplótipos)

C[ATT] ₁₀ AGG	Haplótipo 1 (25%)
T[ATT] ₁₅ GAC	Haplótipo 2 (25%)
T[ATT] ₁₂ AAG	Haplótipo 3 (12,5%)
C[ATT] ₁₅ AAG	Haplótipo 4 (37,5%)

Suponha que o alelo G nesta posição aumente a taxa de crescimento

Variantes formam blocos (haplótipos)

C[ATT] ₁₀ AGG	Haplótipo 1 (25%)
T[ATT] ₁₅ GAC	Haplótipo 2 (25%)
T[ATT] ₁₂ AAG	Haplótipo 3 (12,5%)
C[ATT] ₁₅ AAG	Haplótipo 4 (37,5%)

O alelo G ocorre apenas no haplótipo 2

Variantes formam blocos (haplótipos)

C[ATT]₁₀AG**G** Haplótipo 1 (25%)

T[ATT]₁₅GAC**C** Haplótipo 2 (25%)

T[ATT]₁₂AAG**G** Haplótipo 3 (12,5%)

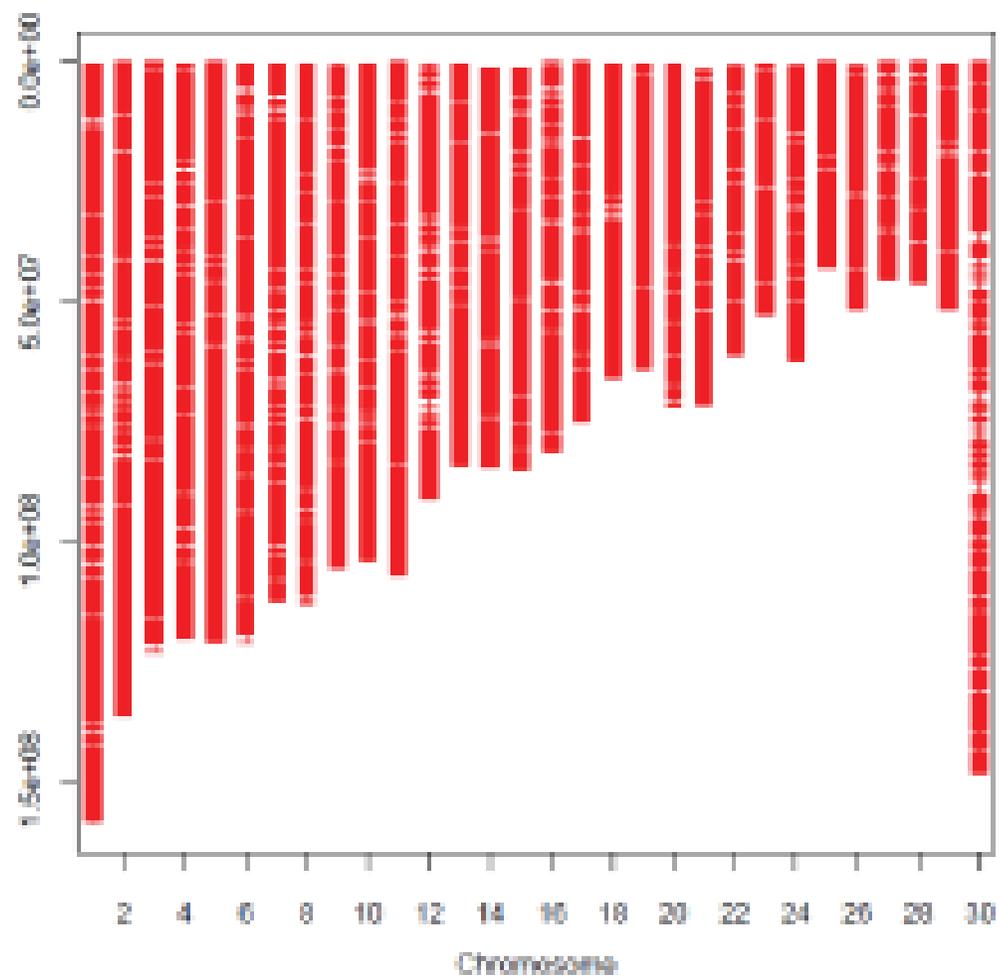
C[ATT]₁₅AA**G** Haplótipo 4 (37,5%)

O alelo C na última posição serve como etiqueta (tag) do alelo G
Dizemos que as duas posições estão em desequilíbrio de ligação

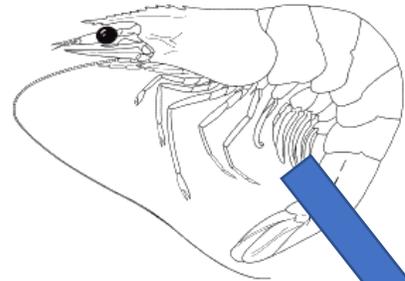


Importante

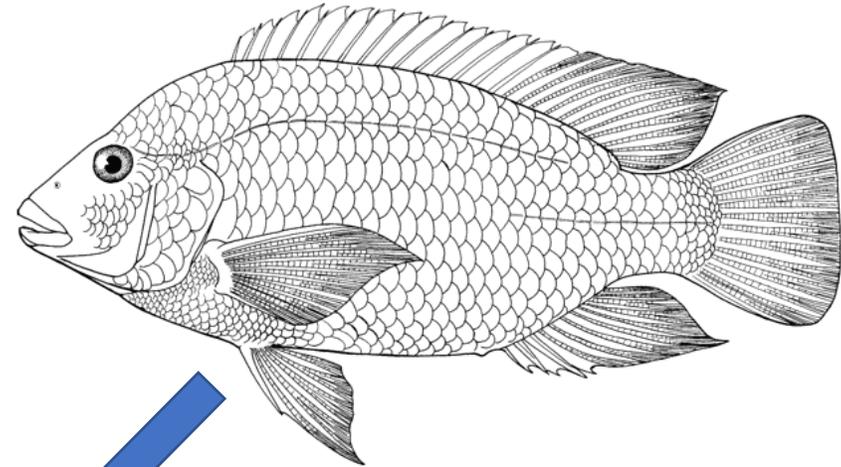
Utilizando-se grande número de SNPs é possível capturar indiretamente a maior parte da variação genômica



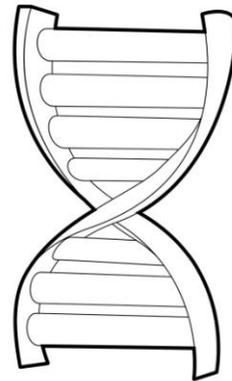
Extração de DNA



Pleópodos



Nadadeira

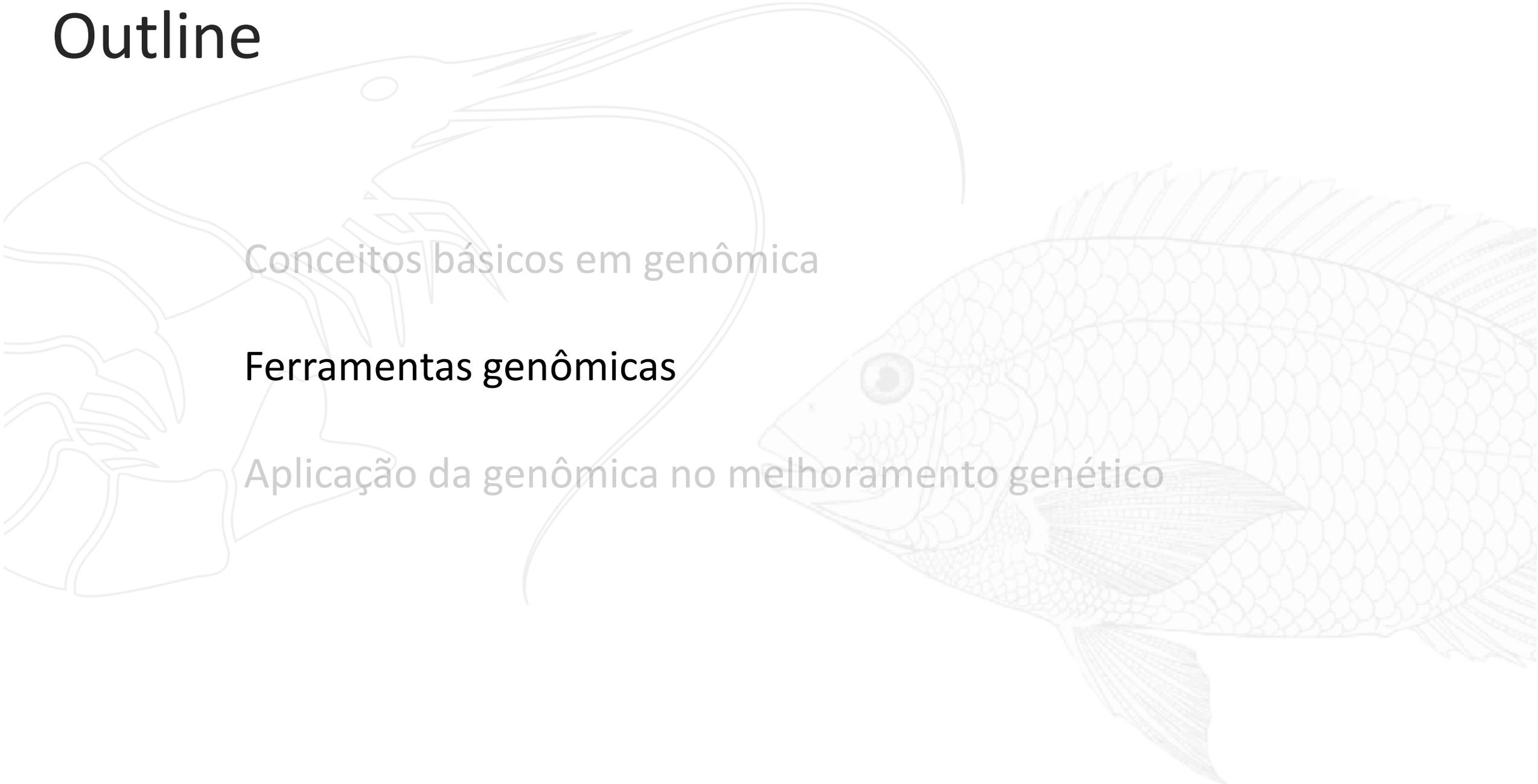


Outline

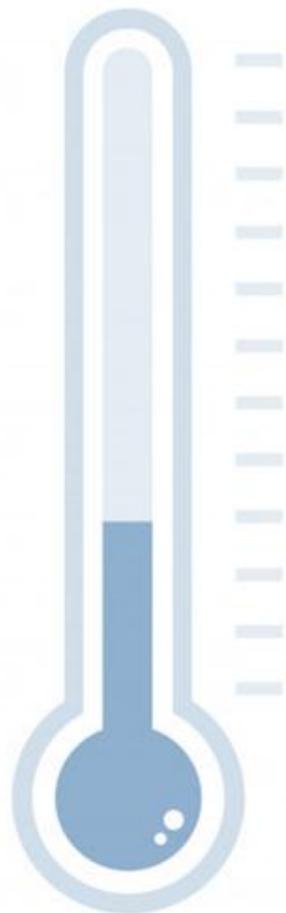
Conceitos básicos em genômica

Ferramentas genômicas

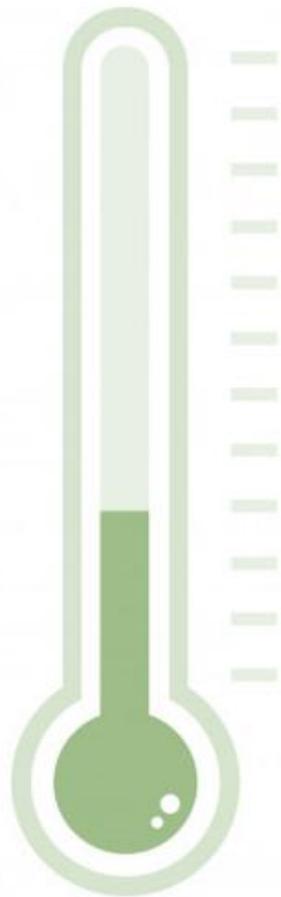
Aplicação da genômica no melhoramento genético



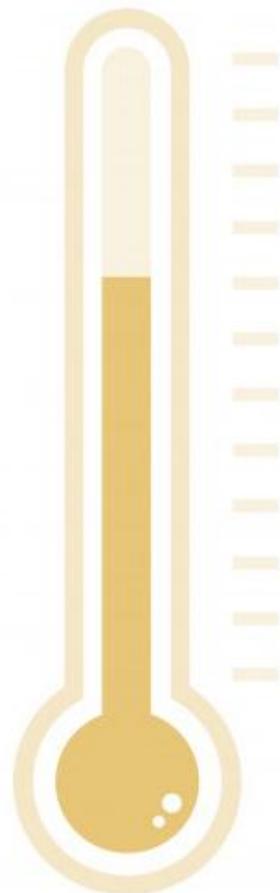
Custo em US\$



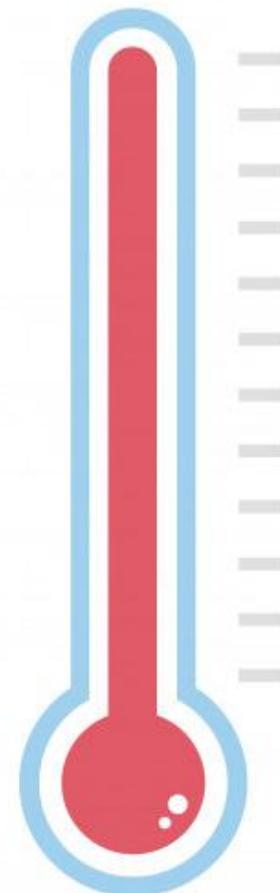
Unidades



Dezenas



Centenas



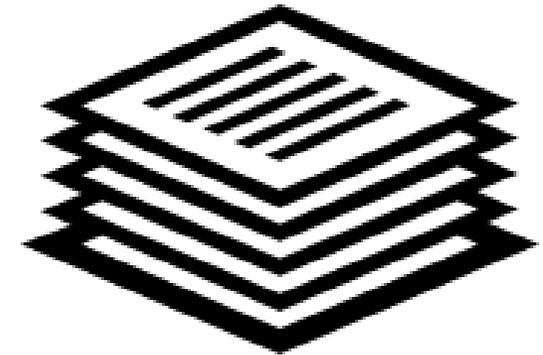
Milhares



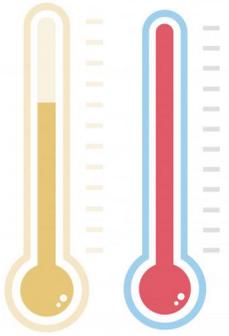
Dados esparsos



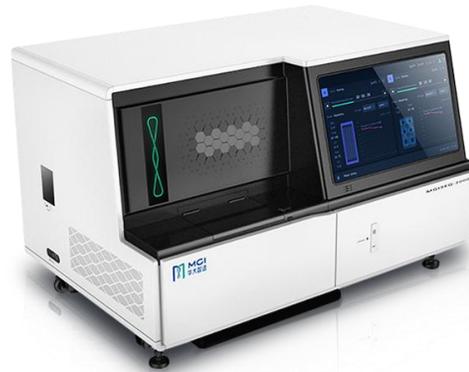
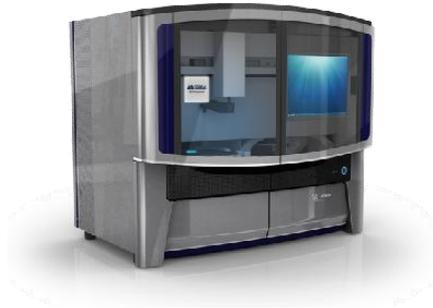
Dados densos



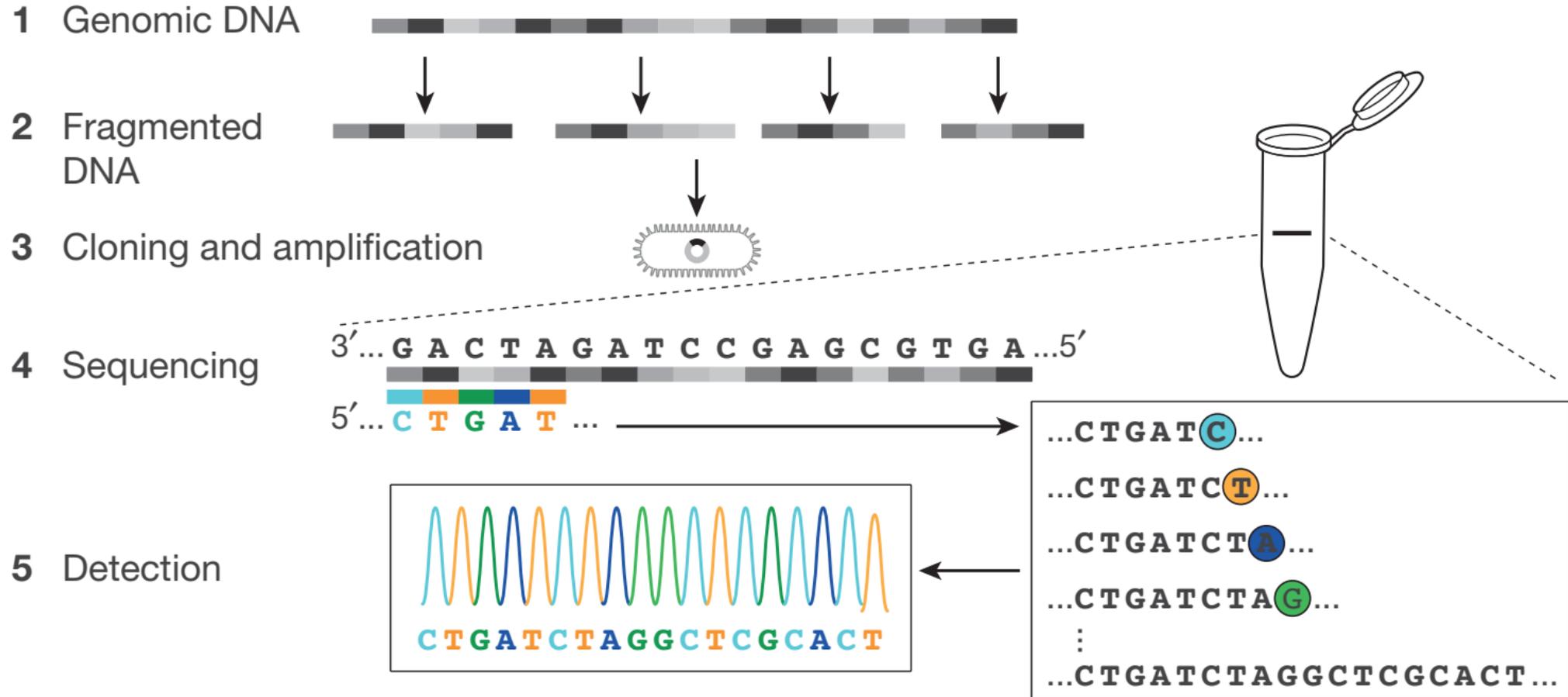
Big data



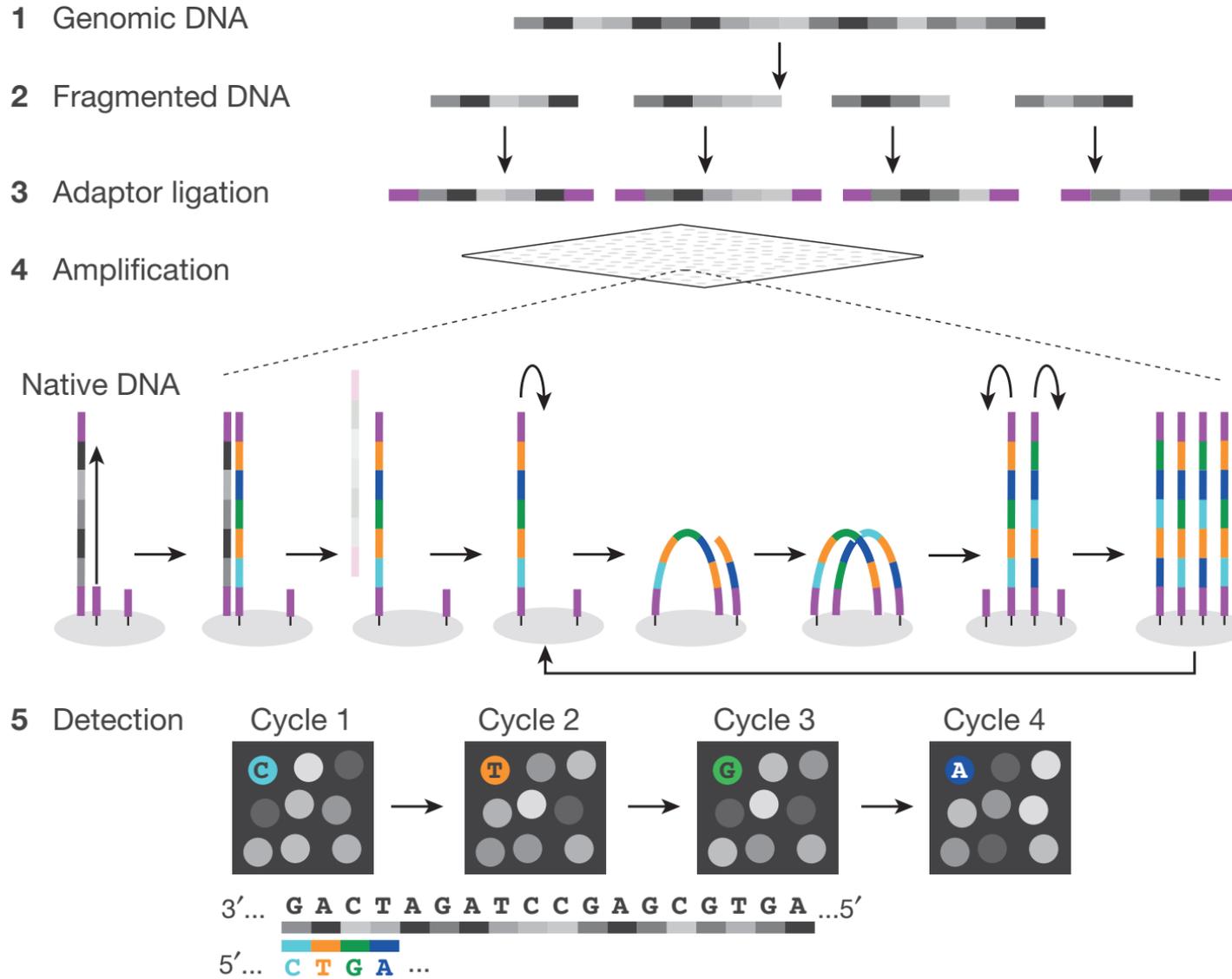
Sequenciamento genômico



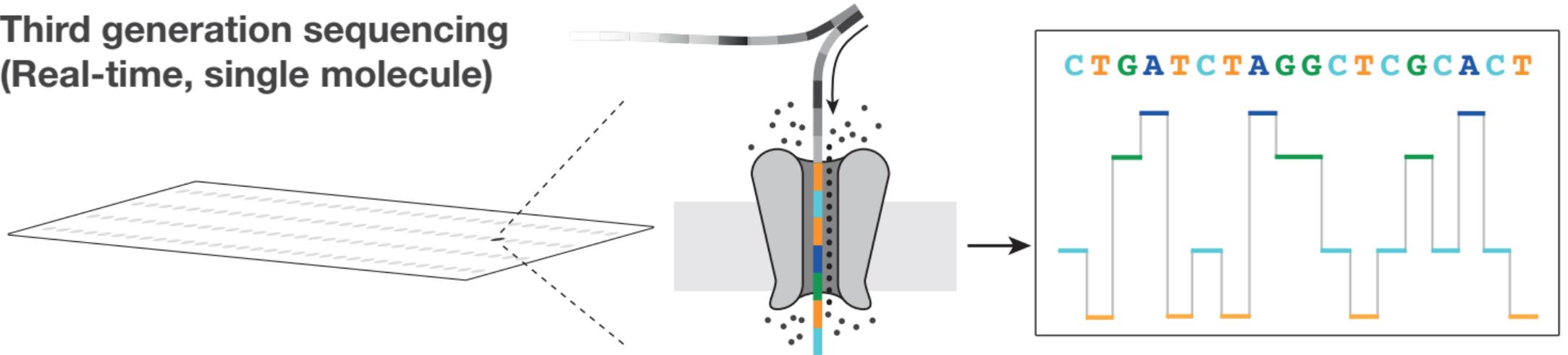
First generation sequencing (Sanger)



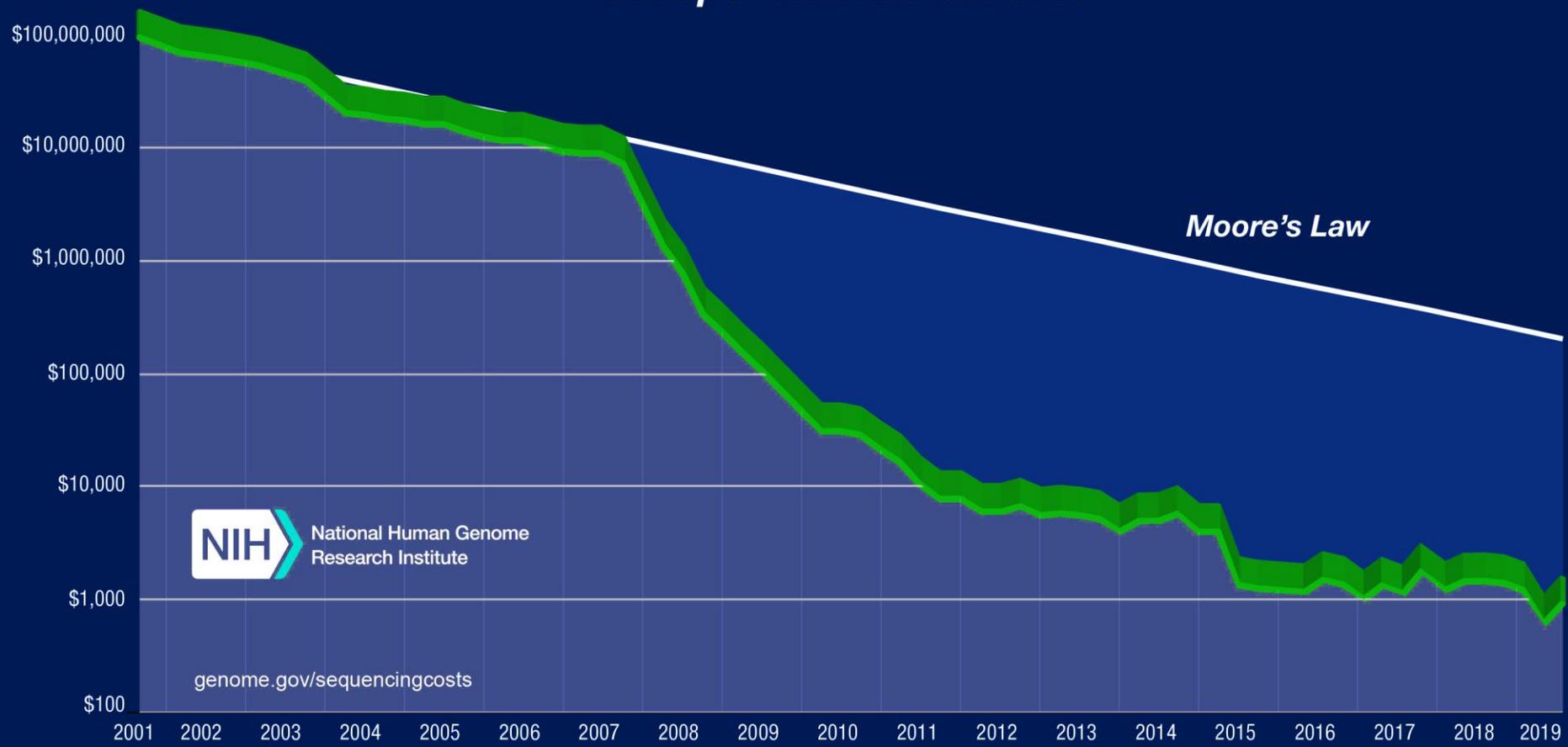
Second generation sequencing (massively parallel)



Third generation sequencing (Real-time, single molecule)



Cost per Human Genome



NIH National Human Genome Research Institute

genome.gov/sequencingcosts

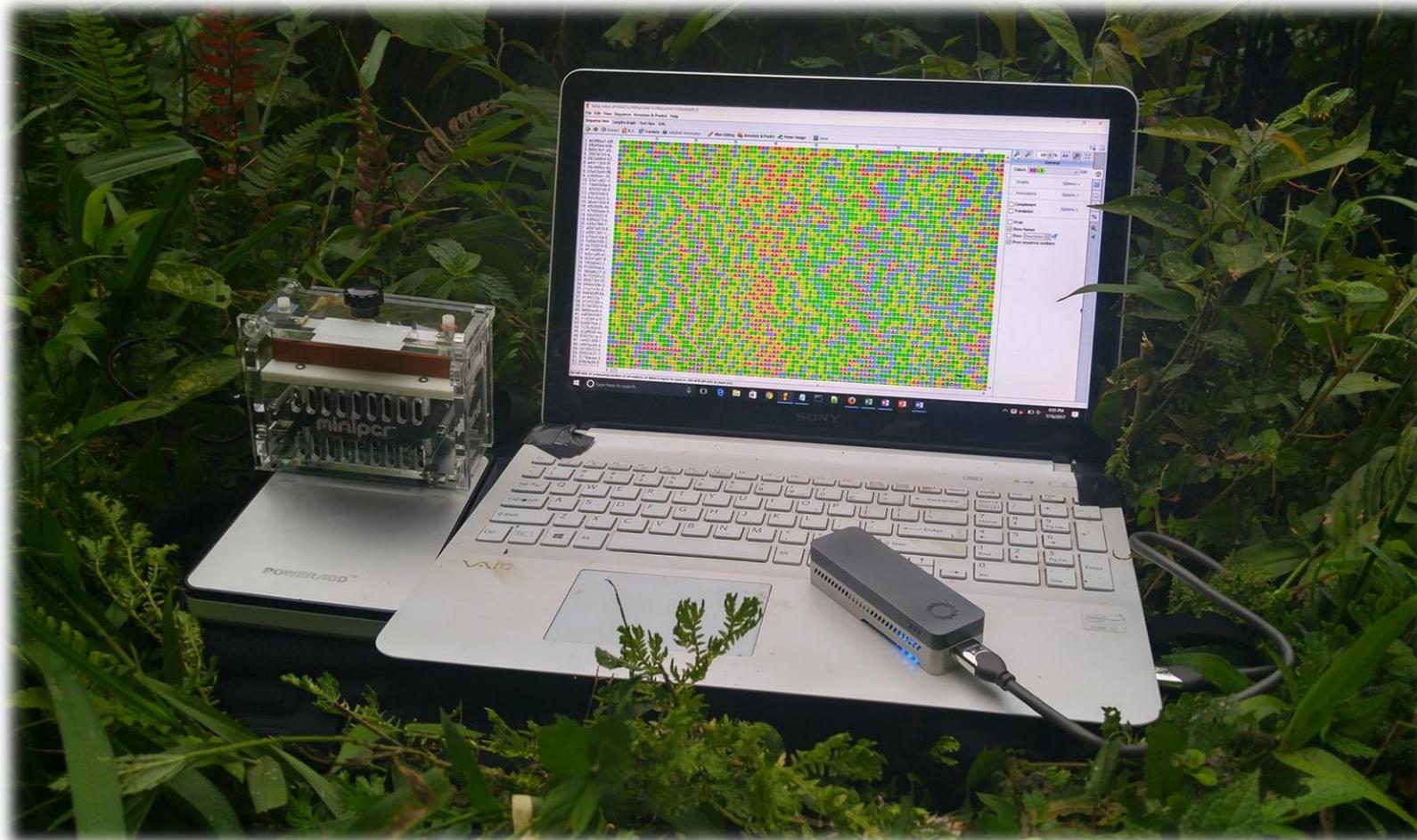


Minion



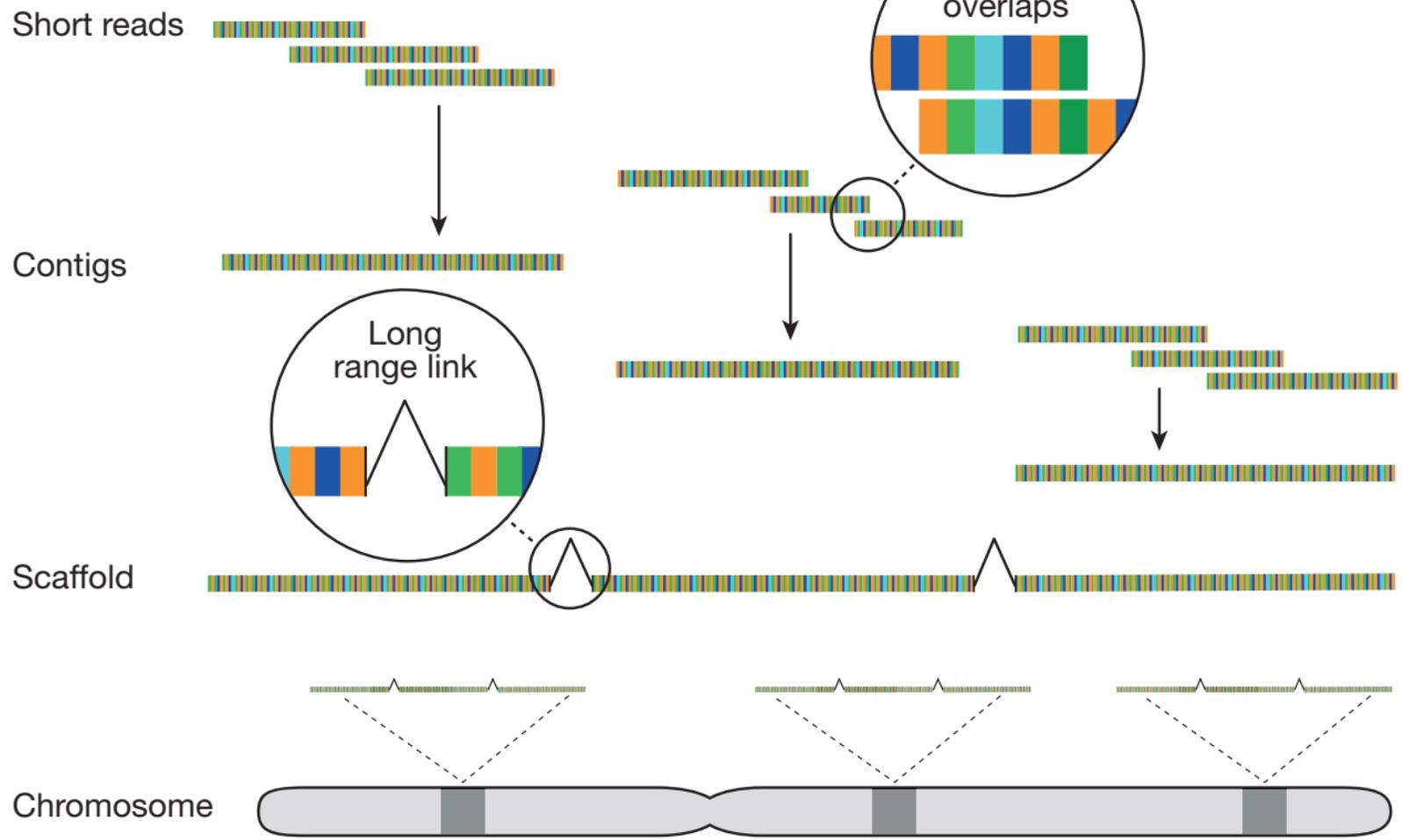
MinION Mk1: portable, real time biological analyses

MinION



Fonte: <http://www.thenextgenscientist.com/home/from-snakes-to-sequences-in-24-hours>

De novo genome assembly



Genoma de referência do camarão



ARTICLE

<https://doi.org/10.1038/s41467-018-08197-4>

OPEN

Penaeid shrimp genome provides insights into benthic adaptation and frequent molting

Xiaojun Zhang^{1,2}, Jianbo Yuan^{1,2}, Yamin Sun³, Shihao Li^{1,2,4}, Yi Gao^{1,2,4}, Yang Yu^{1,2,4}, Chengzhang Liu^{1,2,4}, Quanchao Wang^{1,2,4}, Xinjia Lv^{1,5}, Xiaoxi Zhang^{1,5}, Ka Yan Ma⁶, Xiaobo Wang⁷, Wenchao Lin⁴, Long Wang⁴, Xueli Zhu⁴, Chengsong Zhang^{1,2,4}, Jiquan Zhang^{1,2,4}, Songjun Jin^{1,2,4}, Kuijie Yu^{1,2,4}, Jie Kong⁸, Peng Xu⁹, Jack Chen¹⁰, Hongbin Zhang¹¹, Patrick Sorgeloos¹², Amir Sagi¹³, Acacia Alcivar-Warren¹⁴, Zhanjiang Liu¹⁵, Lei Wang¹⁶, Jue Ruan⁷, Ka Hou Chu⁶, Bin Liu¹⁶, Fuhua Li^{1,2,4} & Jianhai Xiang^{1,2,4}

Total sequence length	1,663,581,301
Total ungapped length	1,618,053,608
Gaps between scaffolds	0
Number of scaffolds	4,683
Scaffold N50	605,555
Scaffold L50	892
Number of contigs	33,020
Contig N50	86,864
Contig L50	5,435
Total number of chromosomes and plasmids	1
Number of component sequences (WGS or clone)	4,683

ASM378908v1

- Organism name: *Penaeus vannamei* (Pacific white shrimp)
- Breed: Kehai No.1
- Sex: male
- BioSample: SAMN08721527; BioProject: PRJNA438564
- Assembly level: Scaffold
- GenBank assembly accession: GCA_003789085.1 (latest)
- RefSeq assembly accession: GCF_003789085.1 (latest)
- WGS Project: QCYY01
- Assembly method: WTDBG v. 1.2
- Genome coverage: 84.0x
- Sequencing technology: PacBio; Illumina HiSeq

Genoma de referência da tilápia

Conte et al. *BMC Genomics* (2017) 18:341
DOI 10.1186/s12864-017-3723-5

BMC Genomics

RESEARCH ARTICLE

Open Access

A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions



Matthew A. Conte¹, William J. Gammerdinger¹, Kerry L. Bartie², David J. Penman² and Thomas D. Kocher^{1*}

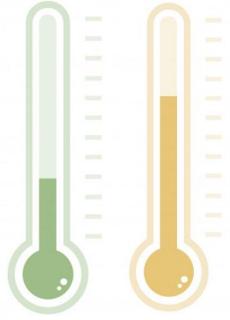
Total sequence length	1,005,681,550
Total ungapped length	1,005,626,550
Gaps between scaffolds	0
Number of scaffolds	2,460
Scaffold N50	38,839,487
Scaffold L50	11
Number of contigs	3,010
Contig N50	2,923,640
Contig L50	96
Total number of chromosomes and plasmids	23
Number of component sequences (WGS or clone)	2,460

O_niloticus_UMD_NMBU

- Organism name: *Oreochromis niloticus* (Nile tilapia)
- Isolate: F11D_XX
- Sex: female
- BioSample: SAMN05859795; BioProject: PRJNA344471
- Assembly level: Chromosome
- GenBank assembly accession: GCA_001858045.3 (latest)
- RefSeq assembly accession: GCF_001858045.2 (latest)
- WGS Project: MKQE02
- Assembly method: Canu v. 1.0
- Genome coverage: 44.0x
- Sequencing technology: PacBio

Re-sequenciamento

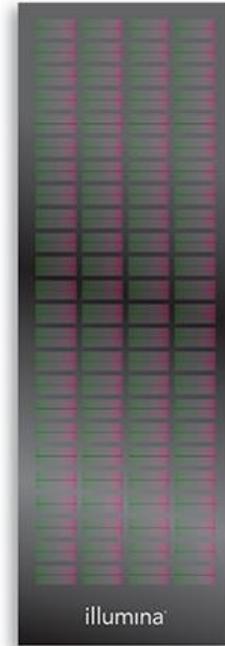




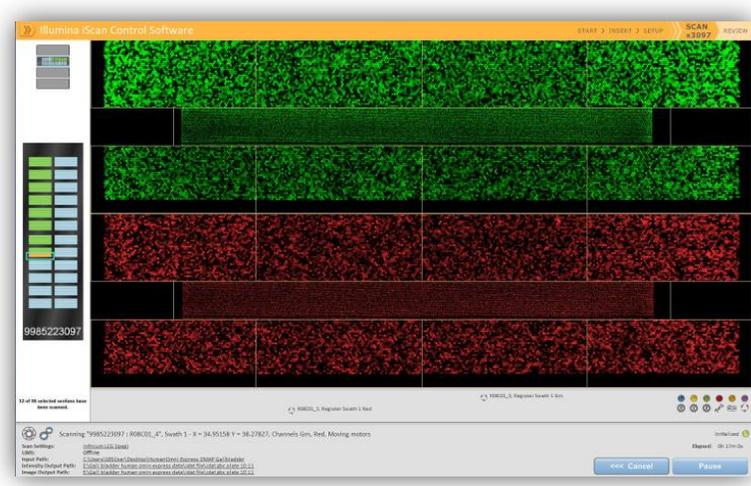
SNP Chips



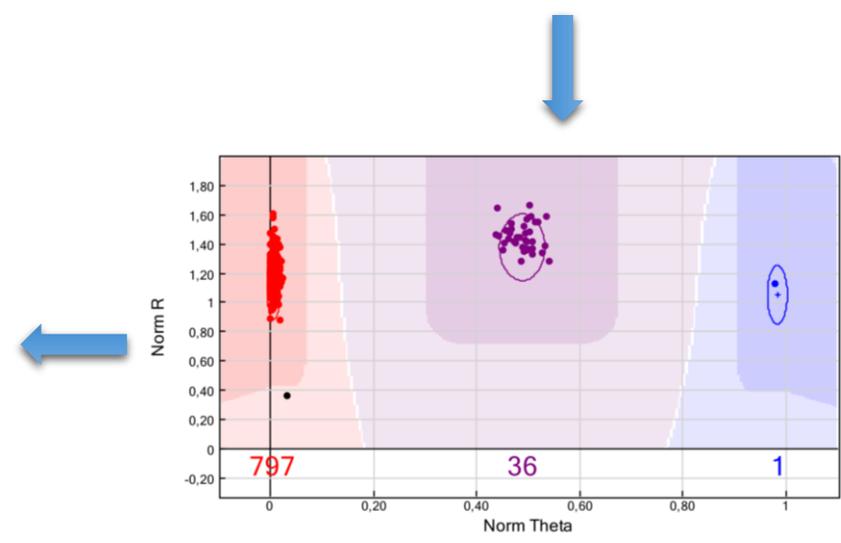
Affymetrix



Illumina

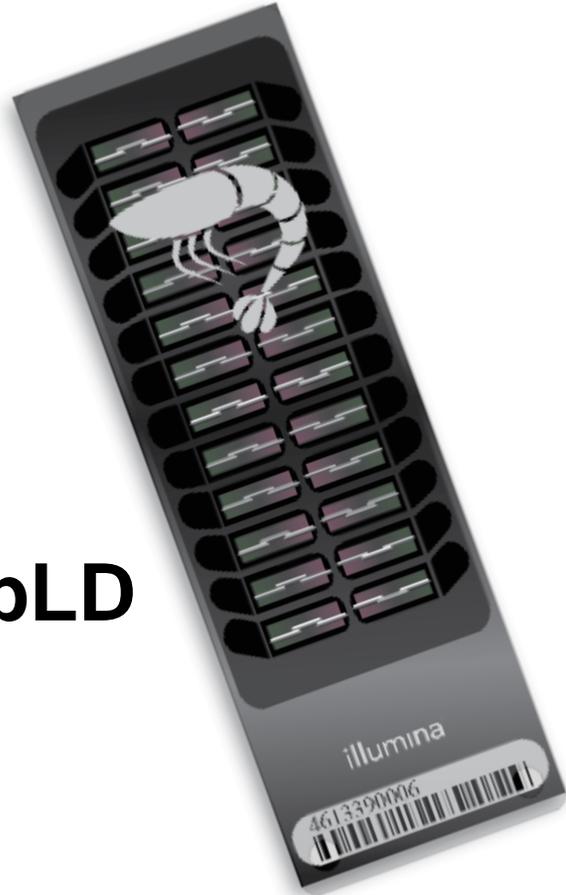


```
[Header]
GSGT Version      1.9.4
Processing Date   02/03/2012 12:32
Content           BovineHD_B.bpm
Num SNPs         777962
Total SNPs       777962
Num Samples      836
Total Samples    836
File             1 of 836
[Data]
SNP Name      Sample ID  Allele1 - Forward  Allele2 - Forward  Allele1 - Top  Allele2 - Top  Allele1 - AB  Allele2 - AB  Log R Ratio  B Allele Freq  GC Score
ARS-BFGL-BAC-10172  ID1      G      G      G      G      B      B      -0.0050  0.9923  0.9374
ARS-BFGL-BAC-1020   ID1      G      G      G      G      B      B      -0.2459  0.9864  0.9568
ARS-BFGL-BAC-10245  ID1      C      C      G      G      B      B      -0.0133  0.9940  0.7996
ARS-BFGL-BAC-10345  ID1      A      A      A      A      A      A      0.1916  0.0000  0.9604
ARS-BFGL-BAC-10365  ID1      G      G      C      C      B      B      -0.3349  0.9874  0.5296
ARS-BFGL-BAC-10591  ID1      A      A      A      A      A      A      0.3282  0.0000  0.8934
ARS-BFGL-BAC-10793  ID1      C      C      G      G      B      B      -0.0418  0.9950  0.9549
ARS-BFGL-BAC-10867  ID1      G      G      C      C      A      A      0.5371  0.0000  0.9592
ARS-BFGL-BAC-10951  ID1      T      T      A      A      A      A      -13.580  0.0117  0.5453
ARS-BFGL-BAC-10952  ID1      A      A      A      A      A      A      0.2831  0.0000  0.9274
ARS-BFGL-BAC-10960  ID1      G      G      G      G      B      B      -0.2537  0.9556  0.9529
ARS-BFGL-BAC-10972  ID1      C      C      G      G      B      B      -0.2088  10.000  0.9195
ARS-BFGL-BAC-10975  ID1      G      G      G      G      B      B      -0.3893  10.000  0.8055
ARS-BFGL-BAC-11000  ID1      T      T      A      A      A      A      0.0202  0.0089  0.9575
ARS-BFGL-BAC-11003  ID1      C      C      G      G      B      B      0.0601  0.9984  0.9177
ARS-BFGL-BAC-11007  ID1      T      T      A      A      A      A      -0.4491  0.0048  0.8796
ARS-BFGL-BAC-11025  ID1      T      T      A      A      A      A      -0.1979  0.0000  0.9281
ARS-BFGL-BAC-11028  ID1      A      A      A      A      A      A      0.1585  0.0000  0.9469
ARS-BFGL-BAC-11034  ID1      C      C      G      G      B      B      -0.0113  0.9785  0.9546
```



SNP Chip do camarão

ShrimpLD



Categoria de conteúdo	Numero de SNPs
Conteúdo validado	
Ciobanu et al 2010	145
Du et al 2009	137
NGS Novas descobertas de SNPs	
Conteúdo mapeado	4390
Conteúdo não mapeado	1,793
Conteúdo Total da Produção	6,465
Conteúdo de Pesquisa Não-Prod	373
Total visível	6,838

SNP Chip da tilápia



ORIGINAL RESEARCH
published: 15 October 2018
doi: 10.3389/fgene.2018.00472

Development and Validation of 58K SNP-Array and High-Density Linkage Map in Nile Tilapia (*O. niloticus*)

Rajesh Joshi^{1†}, Mariann Árnýasi^{1†}, Sigbjørn Lien¹, Hans Magnus Gjøen¹, Alejandro Tola Alvarez² and Matthew Kent¹*

¹ Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway, ² Genomar Genetics AS, Trondheim, Norway

SNPs were identified from whole genome resequencing of 32 individuals from the commercial population of the Genomar strain ...

... high-quality genotype data for 43,588 SNPs.

The GST strain of Nile tilapia used in this study originates from the original GIFT population (Eknath et al., 1993). This strain was selected for growth from generation 1 to 14, growth and filet yield from generation 15 to 19, and growth, yield, and robustness from generation 20.

Novo SNP Chip Trigene/Affymetrix

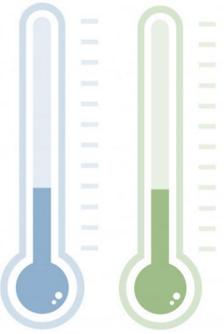
- Variantes identificadas
 - 13 grupos distintos
 - 15.994.791 totais
 - 2.146.867 SNPs bi-alélicos
- Painel incluirá 40.000 SNPs
 - Uniformemente espaçados no genoma
 - Altas frequências nos 19 grupos



Novo SNP Chip Trigene/Affymetrix

- Estratégia para redução de custo
 - Chip multi-espécie
 - 7.000 SNPs de camarão
 - SNPs de outras espécies (ex. Tambaqui e Pacú)





Ensaio baseado em PCR



Microsatélites

Painéis de baixa densidade de SNPs

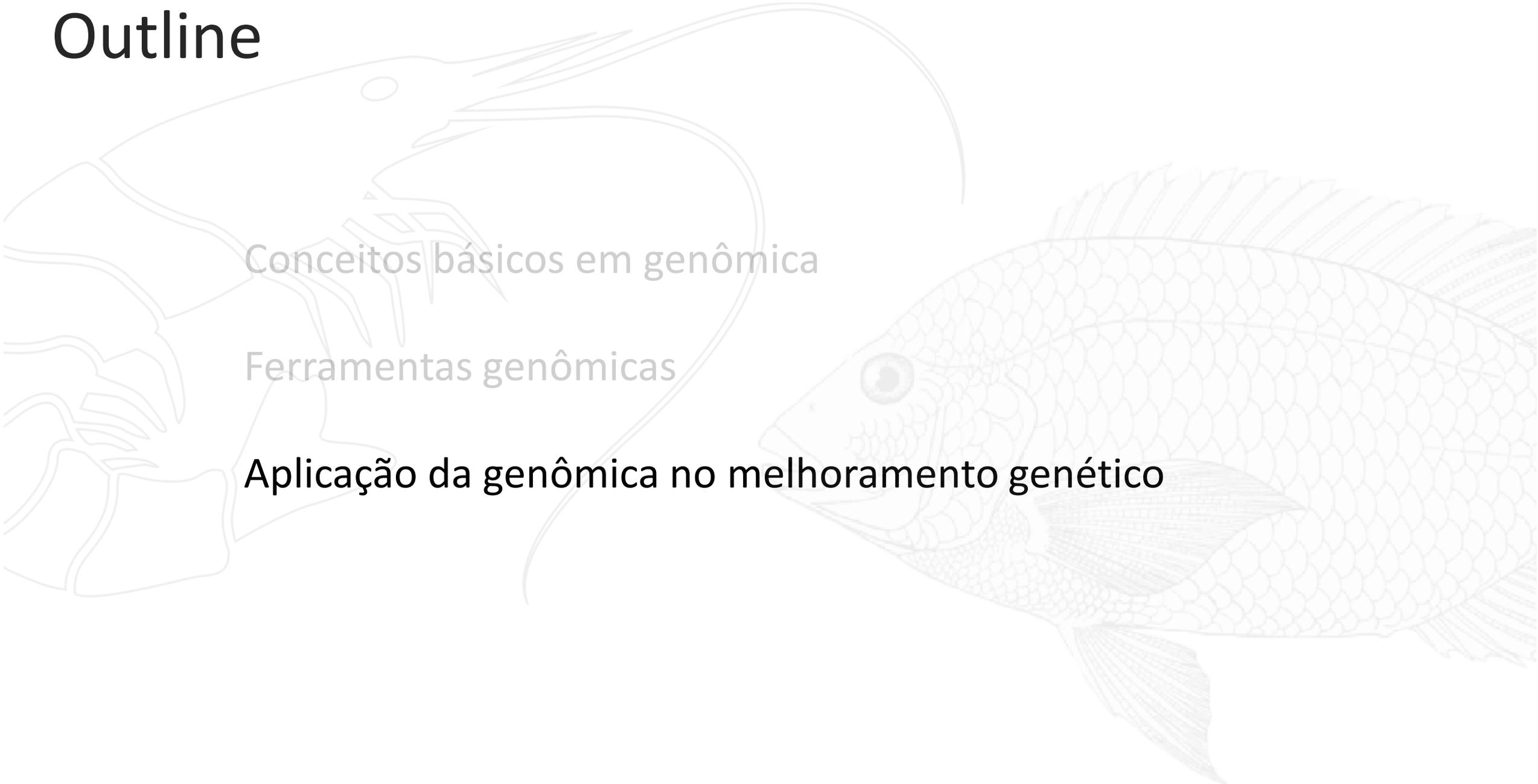


Outline

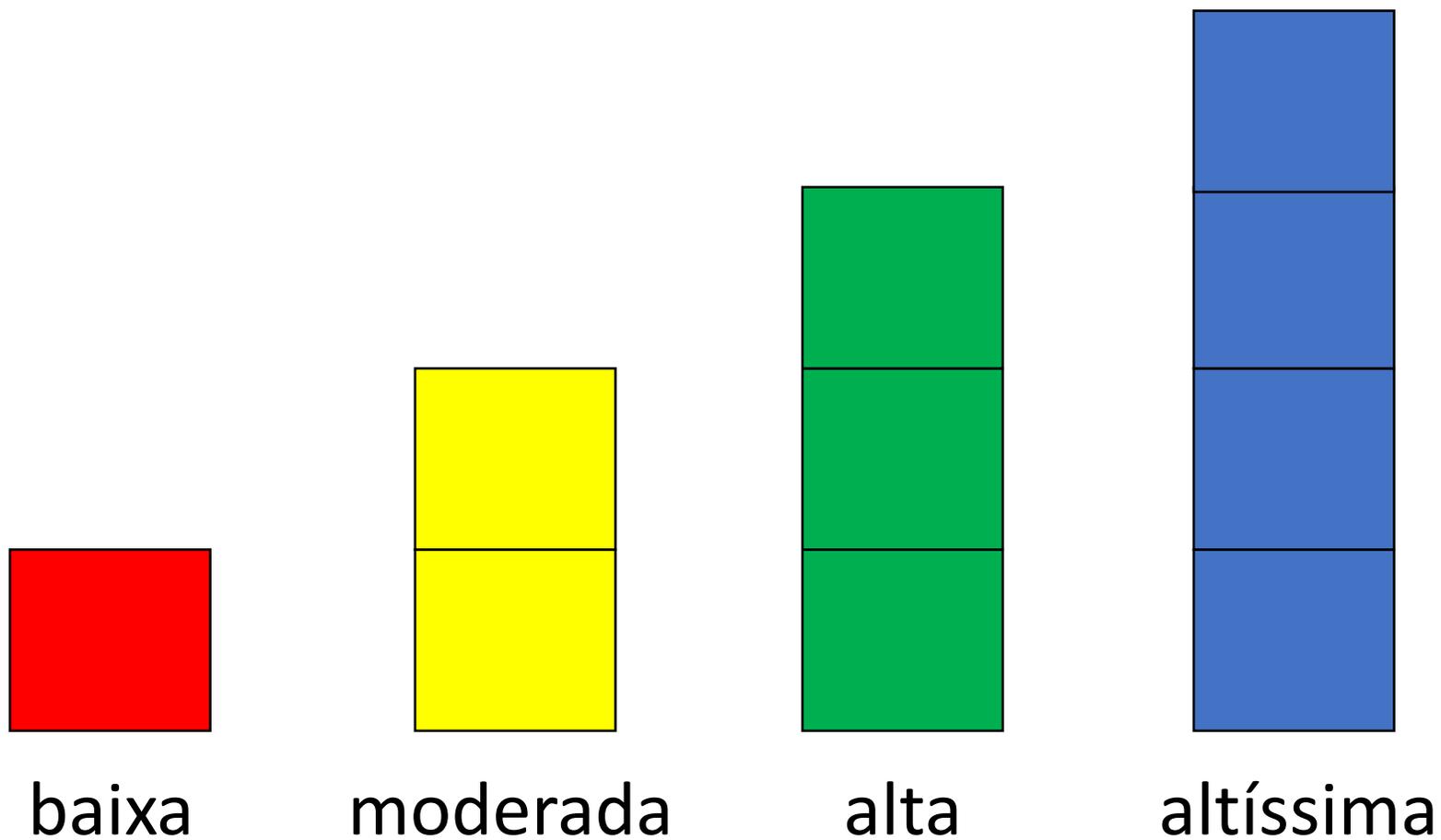
Conceitos básicos em genômica

Ferramentas genômicas

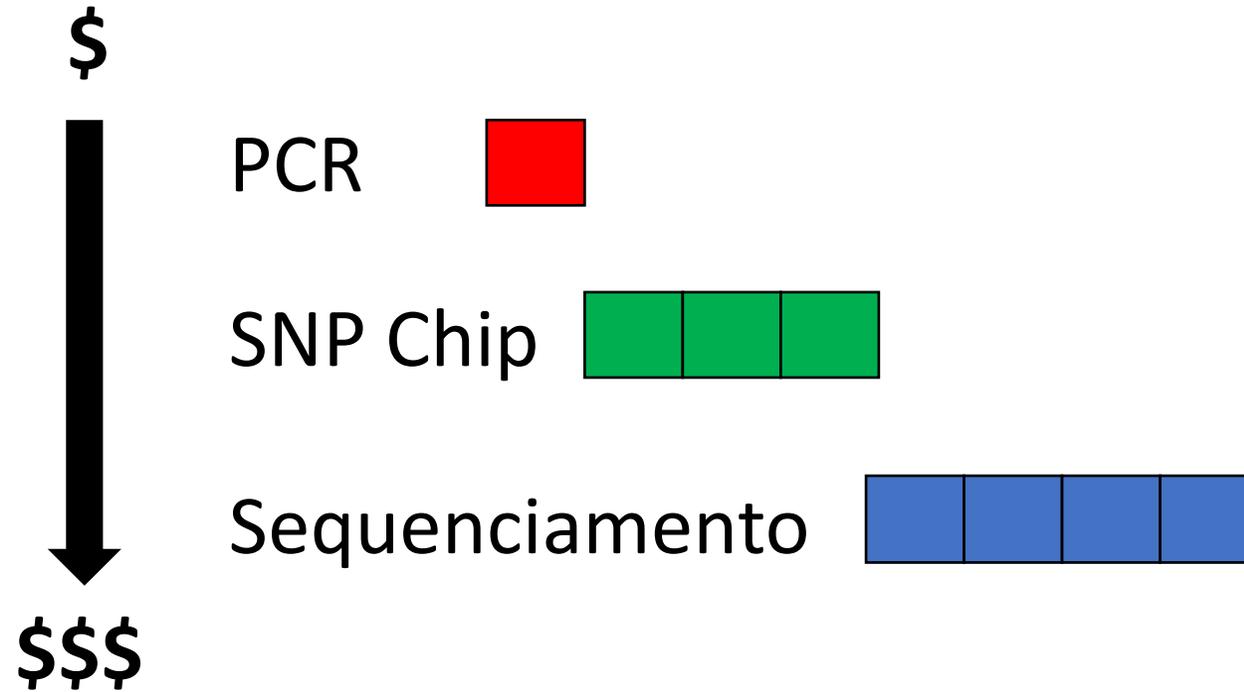
Aplicação da genômica no melhoramento genético



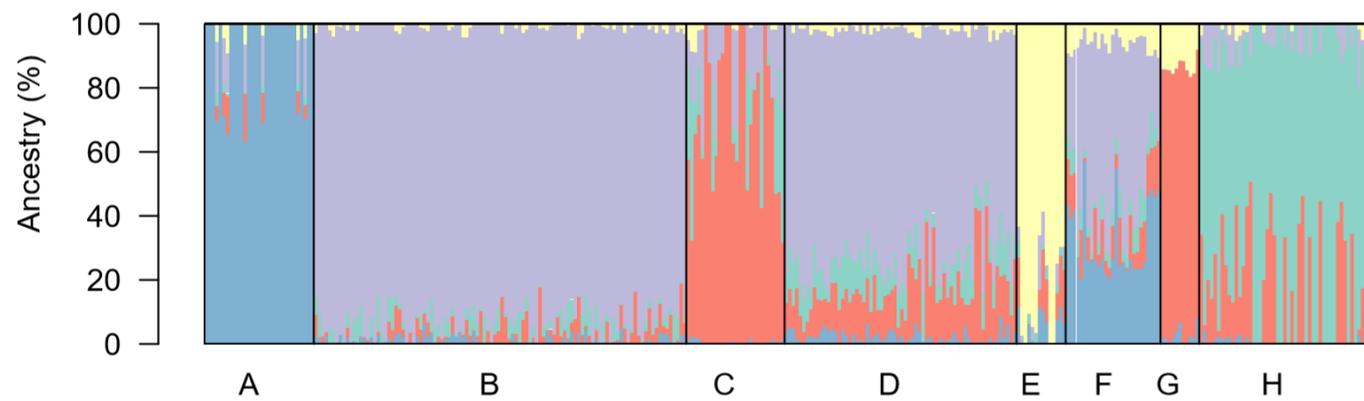
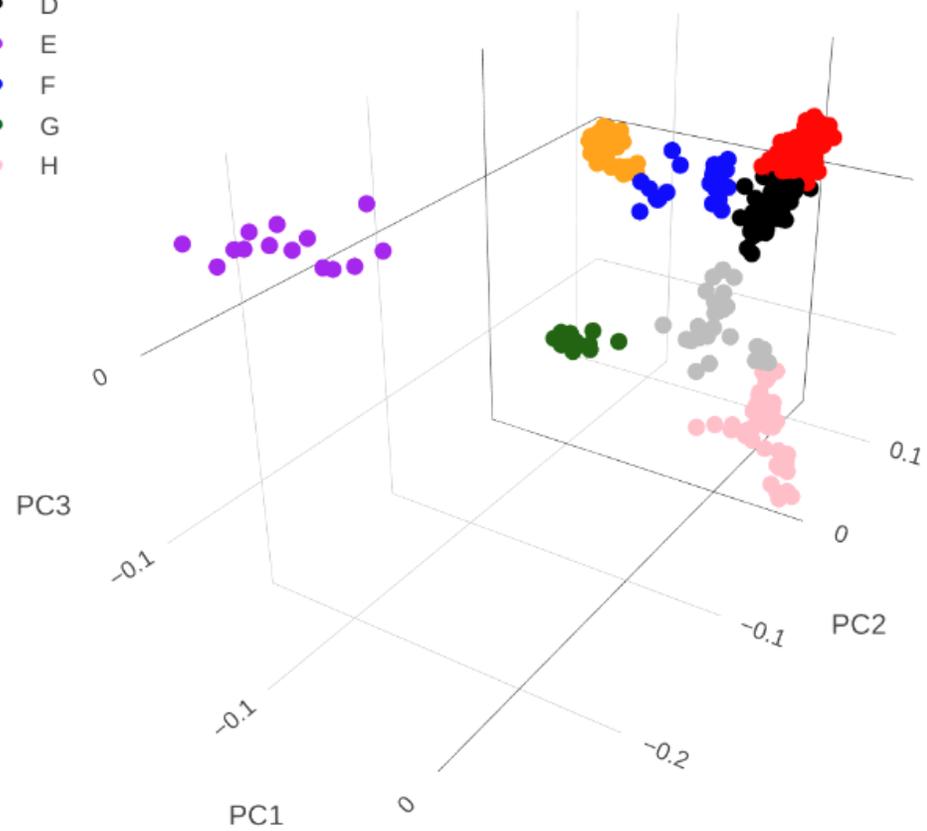
Confiabilidade da análise

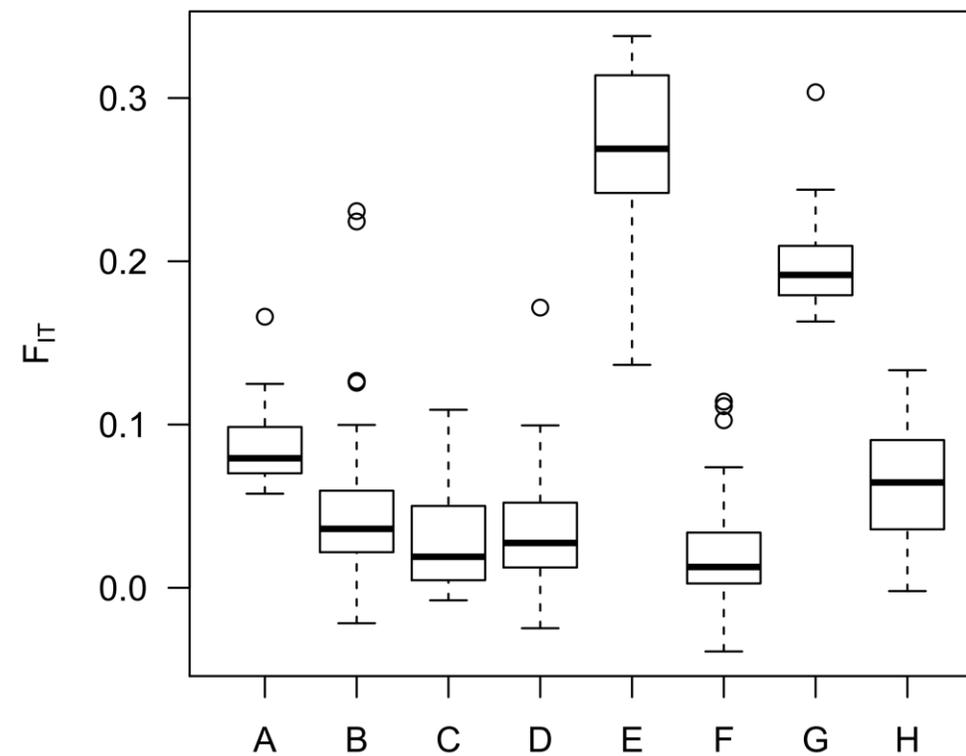
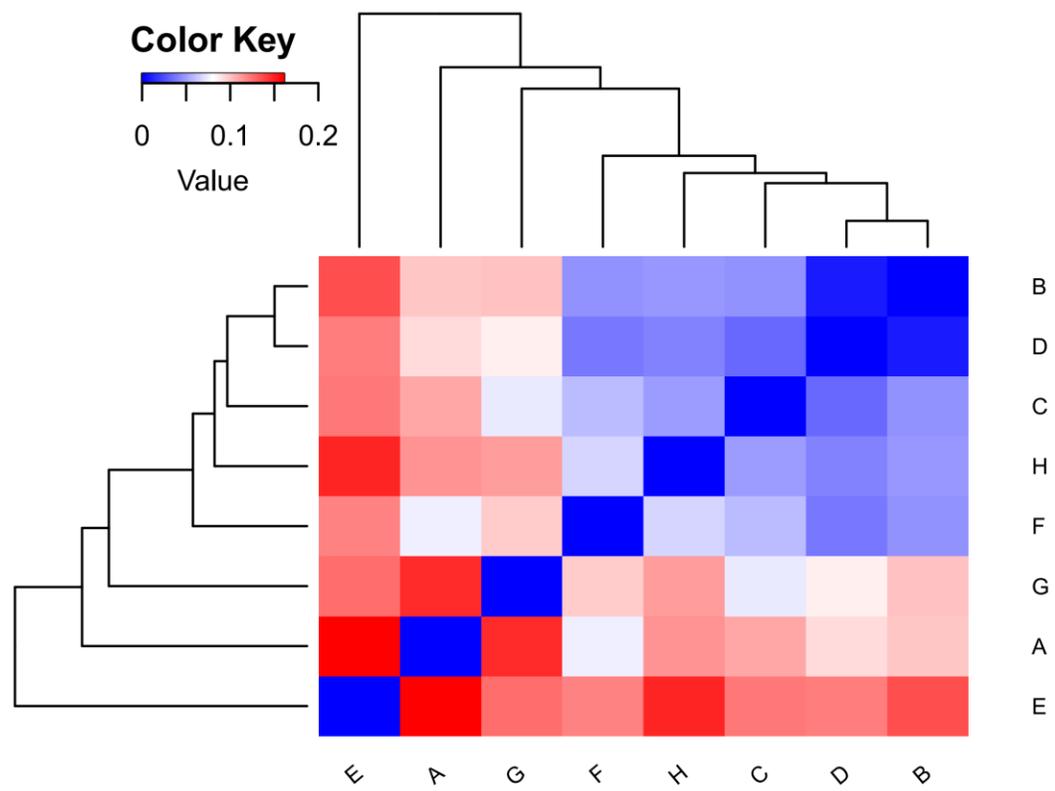


1. Caracterização de linhagens

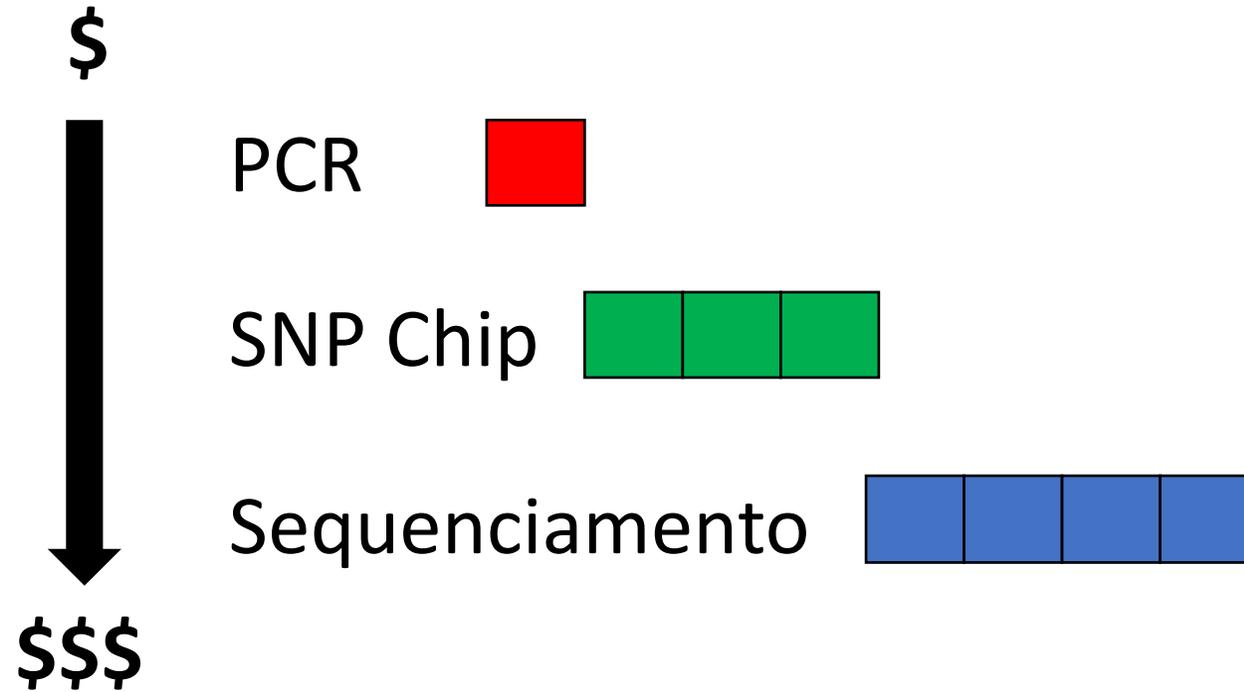


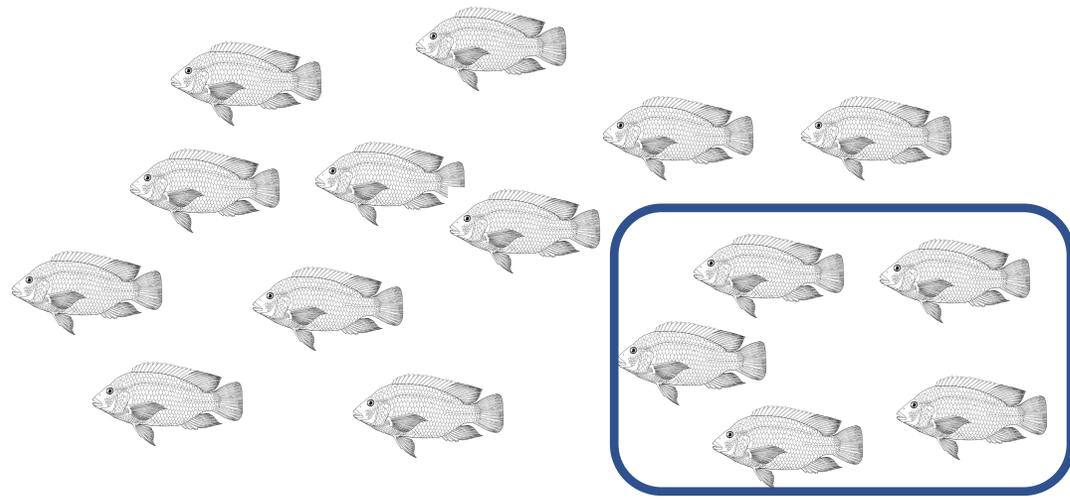
- A
- B
- C
- D
- E
- F
- G
- H



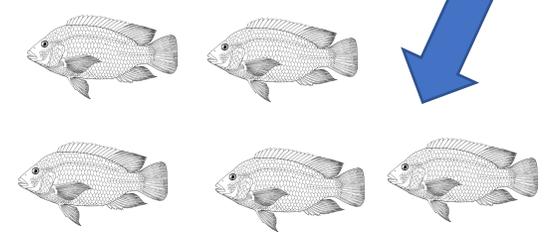


2. Seleção genômica

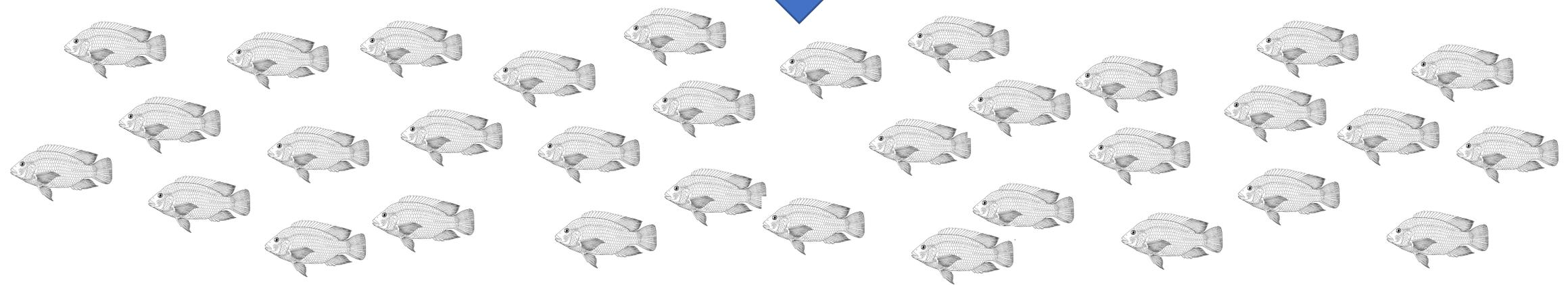




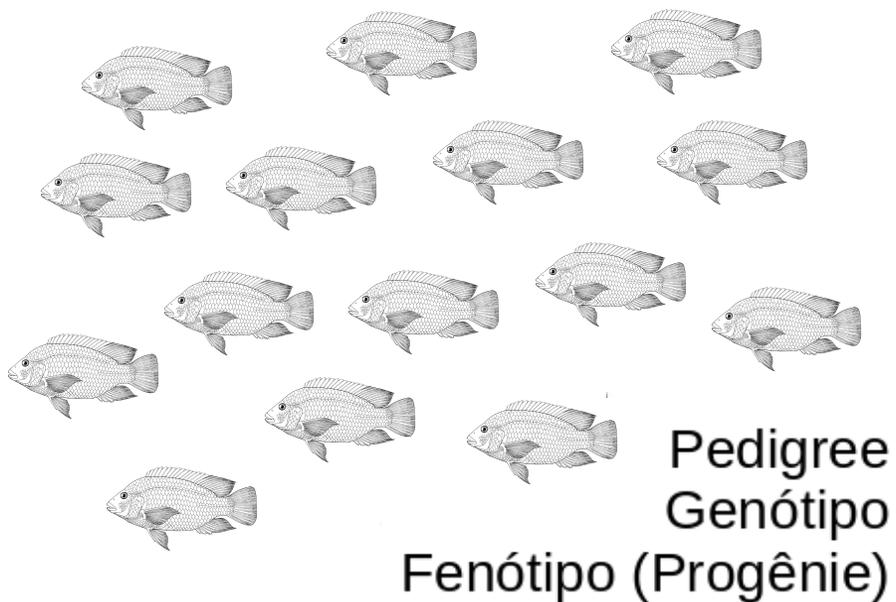
Reprodutores



Reprodução assistida

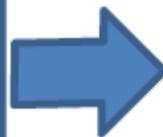


População de referência

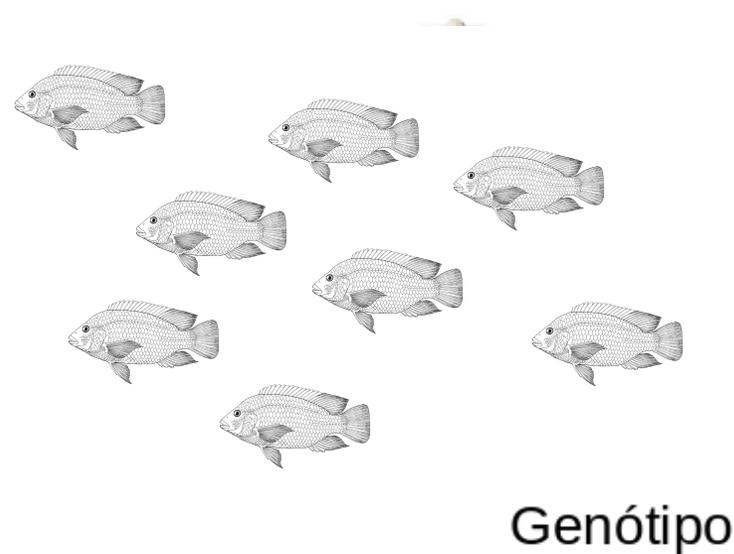


Treinamento

$$\hat{\mathbf{u}}_{\text{ref}} = \mathbf{m}_1 \boldsymbol{\alpha}_1 + \mathbf{m}_2 \boldsymbol{\alpha}_2 + \mathbf{m}_3 \boldsymbol{\alpha}_3 + \dots$$



Candidatos à reprodução



Predição

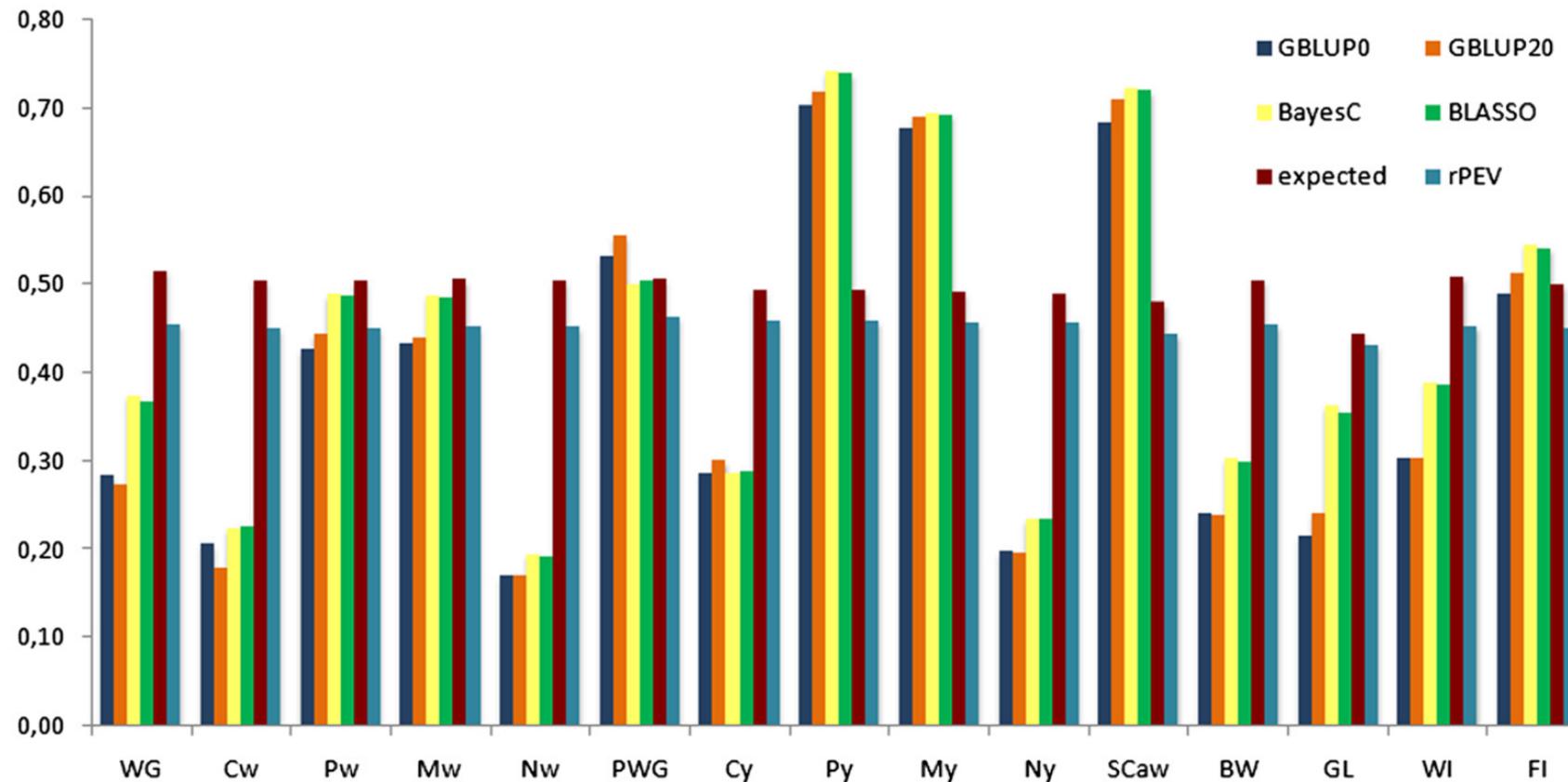
$$\hat{\mathbf{u}}_{\text{test}} = \mathbf{G}_{\text{ref,test}} \mathbf{G}_{\text{ref,ref}}^{-1} \hat{\mathbf{u}}_{\text{ref}}$$

$$\text{Mudança genética} = \frac{\text{Variabilidade} \times \text{Intensidade} \times \text{Acurácia}}{\text{Intervalo entre gerações}}$$

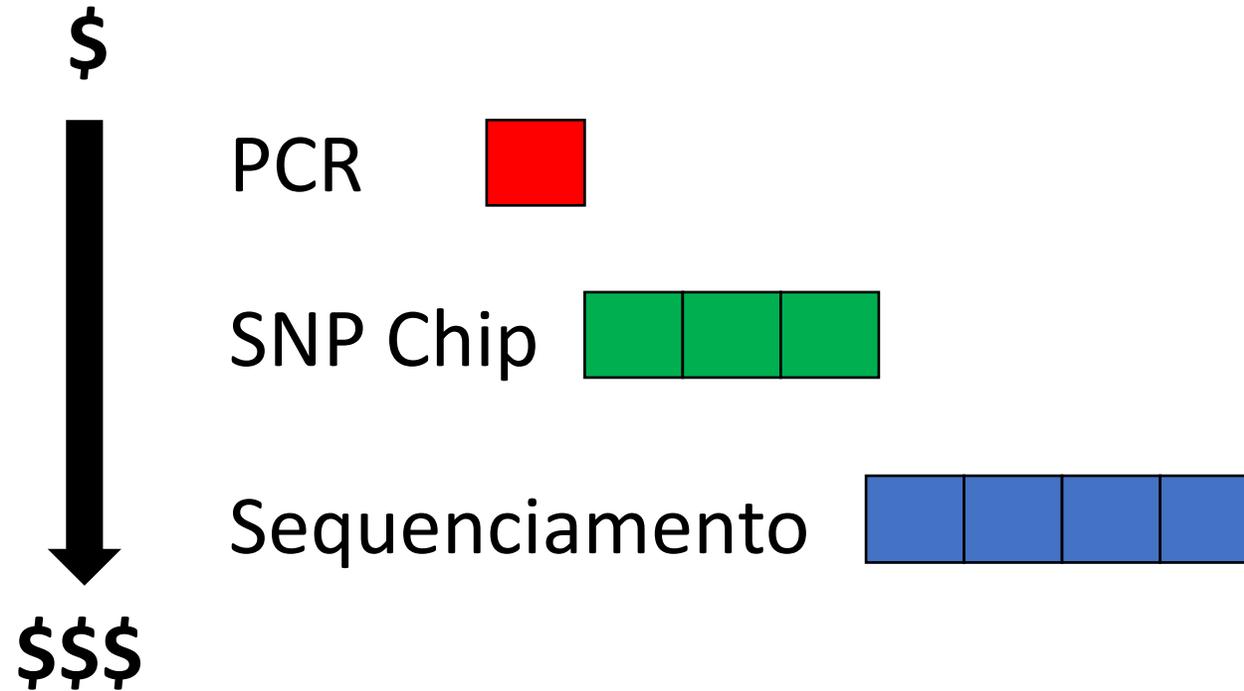
WG
 Cw
 Pw
 Mw
 Nw
 PWG
 Cy
 Py
 My
 Ny
 SCaw
 BW
 GL
 WI
 FI

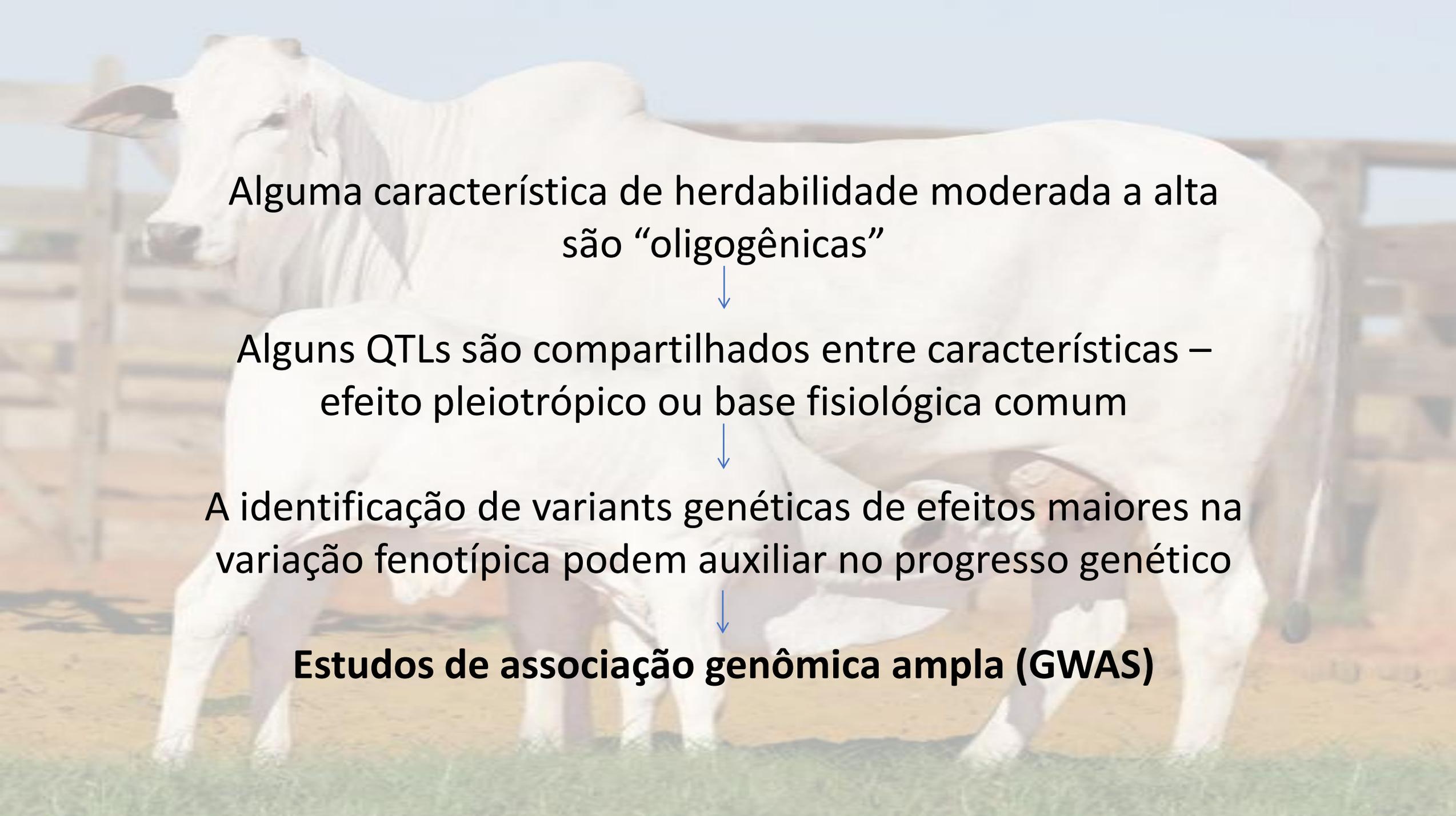
Accuracy of genomic predictions in *Bos indicus* (Nelore) cattle

Haroldo HR Neves^{1*}, Roberto Carvalho^{1,2}, Ana M Pérez O'Brien³, Yuri T Utsunomiya¹, Adriana S do Carmo¹, Flávio S Schenkel⁴, Johann Sölkner³, John C McEwan⁵, Curtis P Van Tassel⁶, John B Cole⁷, Marcos VGB da Silva⁸, Sandra A Queiroz¹, Tad S Sonstegard⁶ and José Fernando García⁹



3. Associação fenótipo-genótipo





Alguma característica de herdabilidade moderada a alta
são “oligogênicas”



Alguns QTLs são compartilhados entre características –
efeito pleiotrópico ou base fisiológica comum



A identificação de variants genéticas de efeitos maiores na
variação fenotípica podem auxiliar no progresso genético

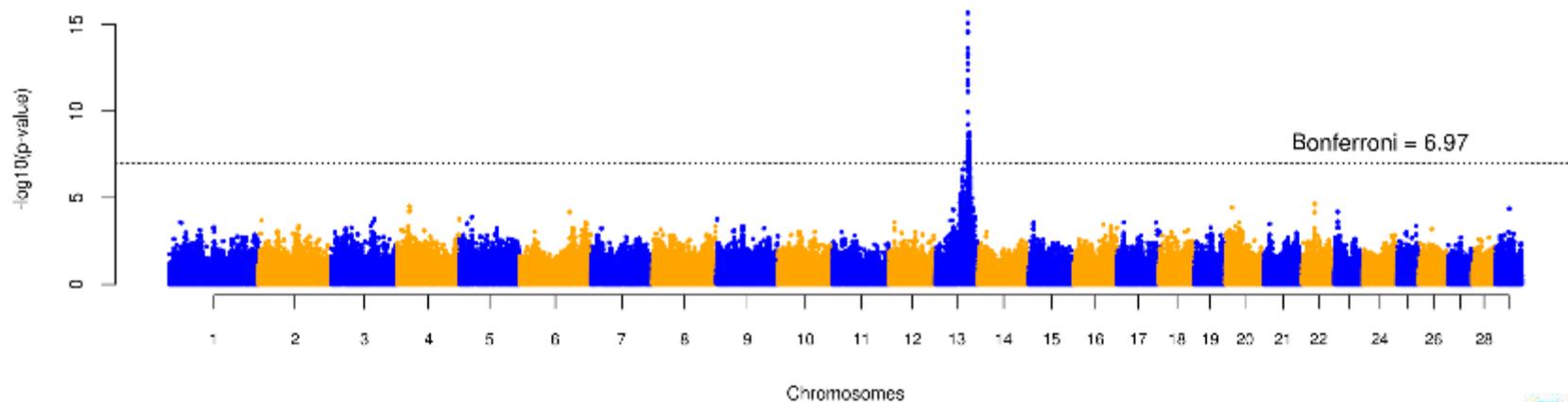


Estudos de associação genômica ampla (GWAS)



N = 131

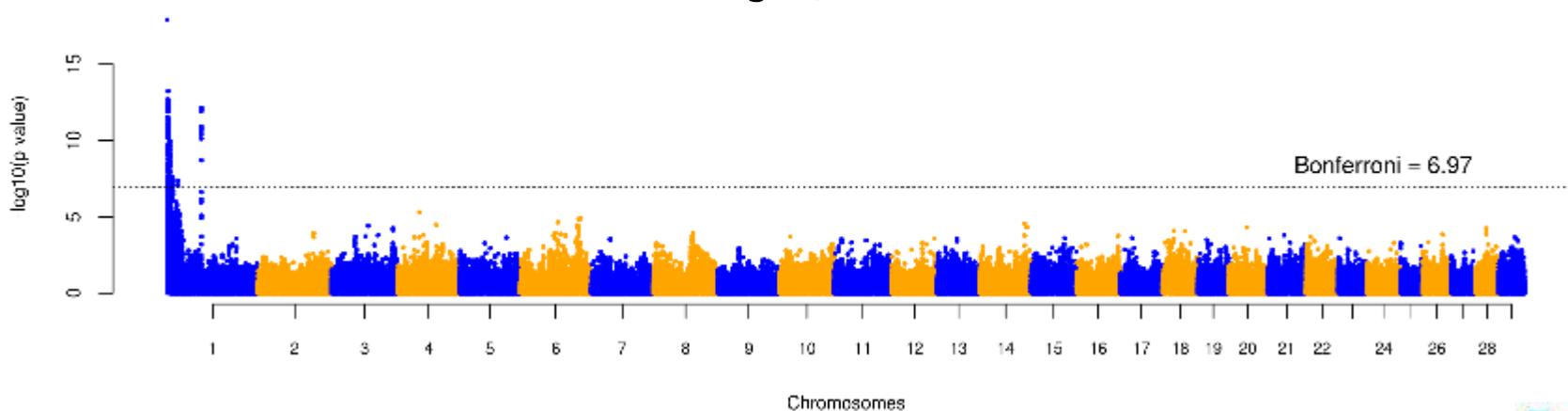
Cor da pelagem





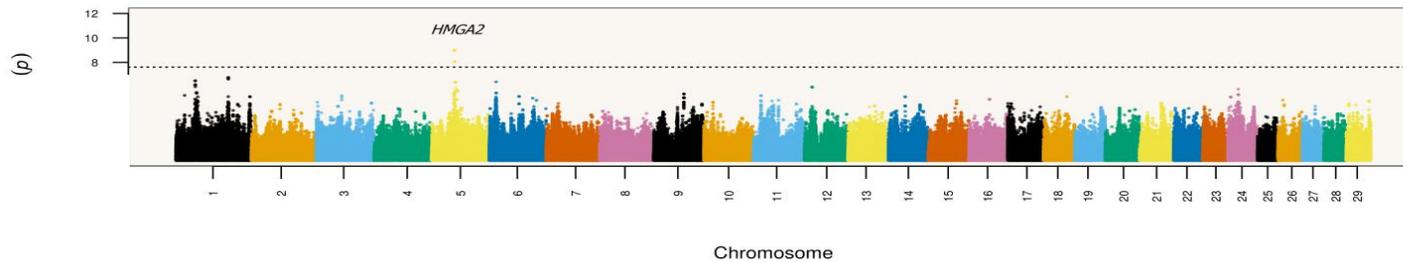
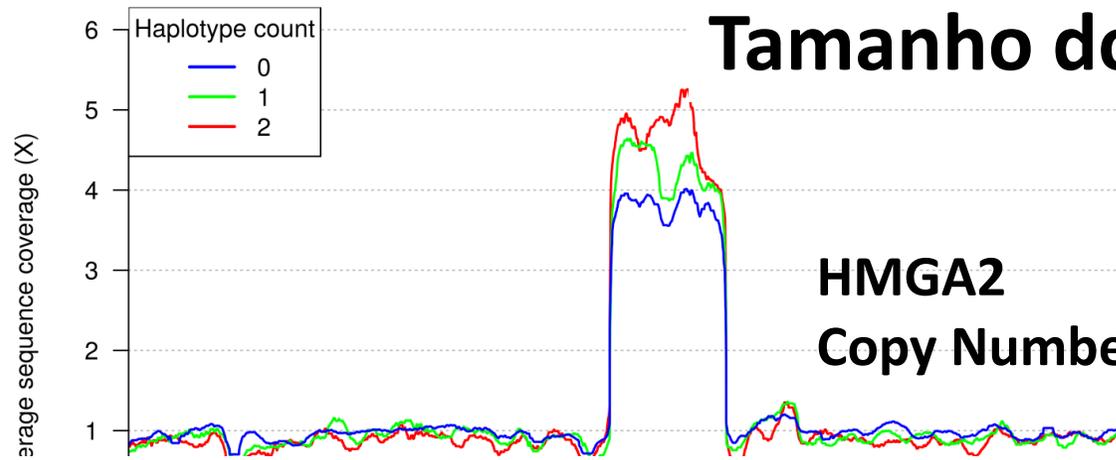
N = 401

Presença/Ausência de chifres





N = 2.231





GDF8 - miostatina

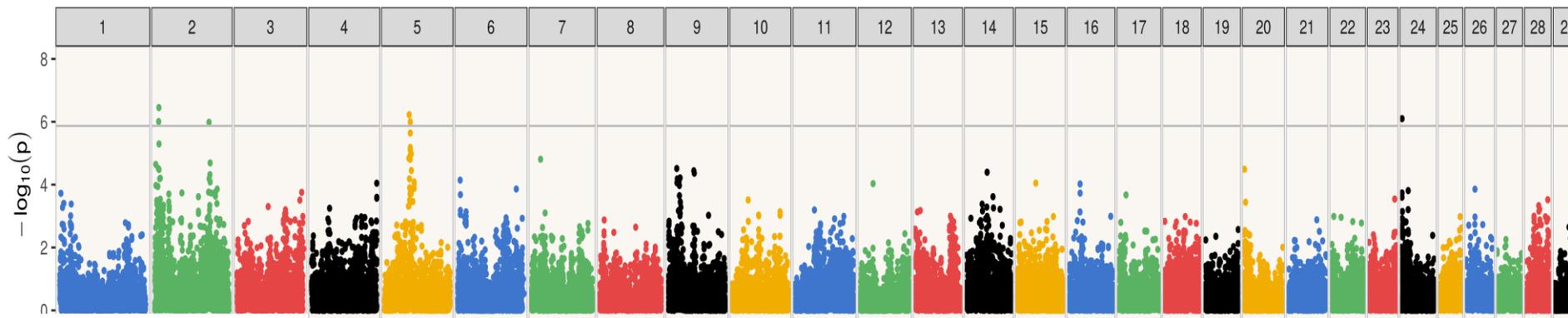
Unknown

HMGA2

Gado Senepol

N = 1.800

Unknown



Marmoreio



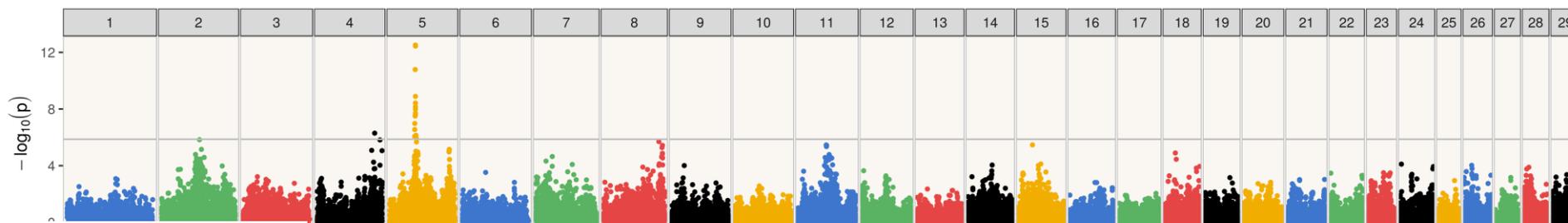


CASP2

Gado Senepol

N = 1.800

HMGA2



Escore de população folicular ovariana





**Precisamos de sistemas
robustos de coleta de
fenótipos!**

Genômica é prata...



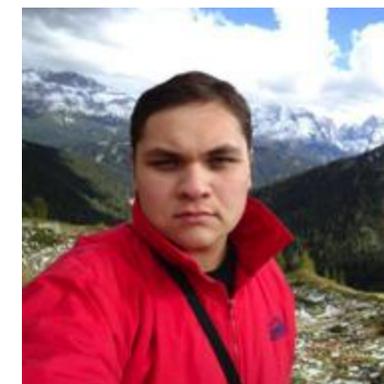
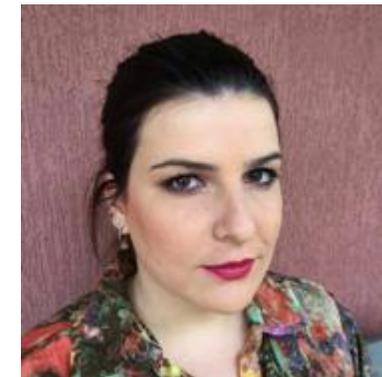
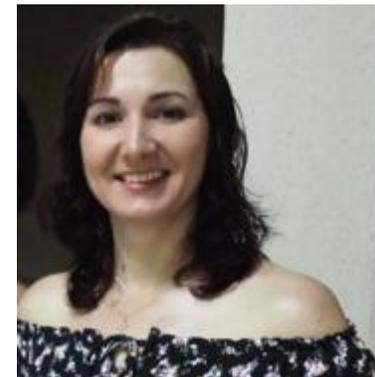
Fenótipo é Ouro!!!!





Nossa equipe

- João L Campos, MSc
- Eduardo A Ono, MSc
- José F Garcia, PhD
- Adam TH Utsunomiya, PhD
- Lucas A Da Silva, BSc
- Marco Milanesi, PhD
- Silvana de C Paulan, PhD
- Rafaela BPT Torrecilha, PhD
- Yuri T Utsunomiya, PhD





Obrigado!

yuri.utsunomiya@unesp.br
contato@agropartners.com.br