Zexin Jiao[1], Robert Mukiibi[1], Manu Kumar Gundappa[1], James Prendergast[1]
Luca Bargelloni[2], Diego Robledo[1], Daniel J. Macqueen[1]

1 The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, UK
2 Department of Comparative Biomedicine and Food Science, University of Padova, Padova, Italy

Email: jiaozexin@outlook.com
z.jiao-3@sms.ed.ac.uk

The structural variation landscape in the European seabass (*Dicentrarchus labrax*) genome and its potential role in disease resistance

Introduction

- **Structural variants (SVs)** are genetic polymorphisms that affect >50bp of sequence, including deletions, insertions, inversions, duplications, and translocations (Fig. 1).
- **European seabass (*Dicentrarchus labrax*)** (Fig. 2) is a high-value aquaculture species, and one of the key infectious diseases affecting seabass aquaculture is viral nervous necrosis (**VNN**).
- Quantitative trait loci (**QTL**) for VNN resistance have been documented across diverse farmed populations based on genome-wide association studies (**GWAS**) using SNPs.
- SVs have a major impact which are more commonly causative functional variants than SNPs. The **lack of SV** research prompted us to develop a study about SVs and their potential roles in key aquaculture species such as **seabass and salmonid**.

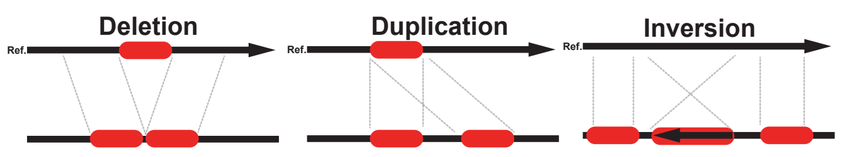

Fig. 1 Major classes of structural variants

Fig. 2 European seabass (*Dicentrarchus labrax*) (Image source: NCBI)

Objectives

- Characterize the **SV landscape** in seabass using whole genome sequencing data from n=90 samples (representing 25 families including parents and offspring).
- Annotate high confidence SVs and understand their **impact** on genes and **functional feature** across the genome.
- **Impute SVs** based on SNP genotypes for 990 offspring using parent genotypes to perform **GWAS for VNN** resistance using SVs.

GWAS analysis using SVs

- Host resistance to VNN was defined as **binary survival (BS)** and as **days to death (DD)**.
- **SV-based heritability estimates** for VNN resistance ranged between 0.3688 ± 0.0568 (BS) and 0.4147 ± 0.0566 (DD), which were consistent with SNP-based heritability estimates.
- **Estimated inflation factors (λ)** for BS and DD were 1.10 and 1.08.
- 108 (BS) and 122 (DD) SVs exceeded genome-wide significance in **a single QTL region** matching to previous work based on SNPs (Fig 6).

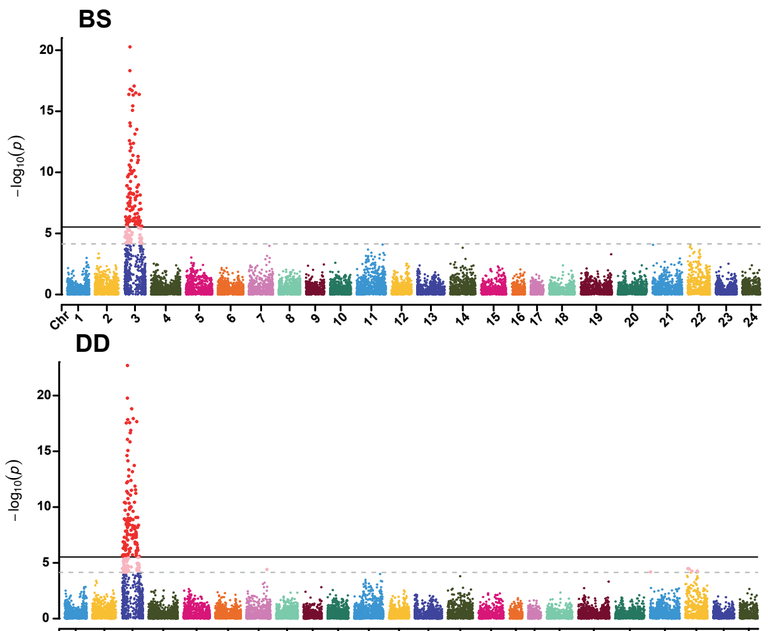

Fig. 6 GWAS analysis for VNN resistance trait based on SV

SV detection and imputation pipeline

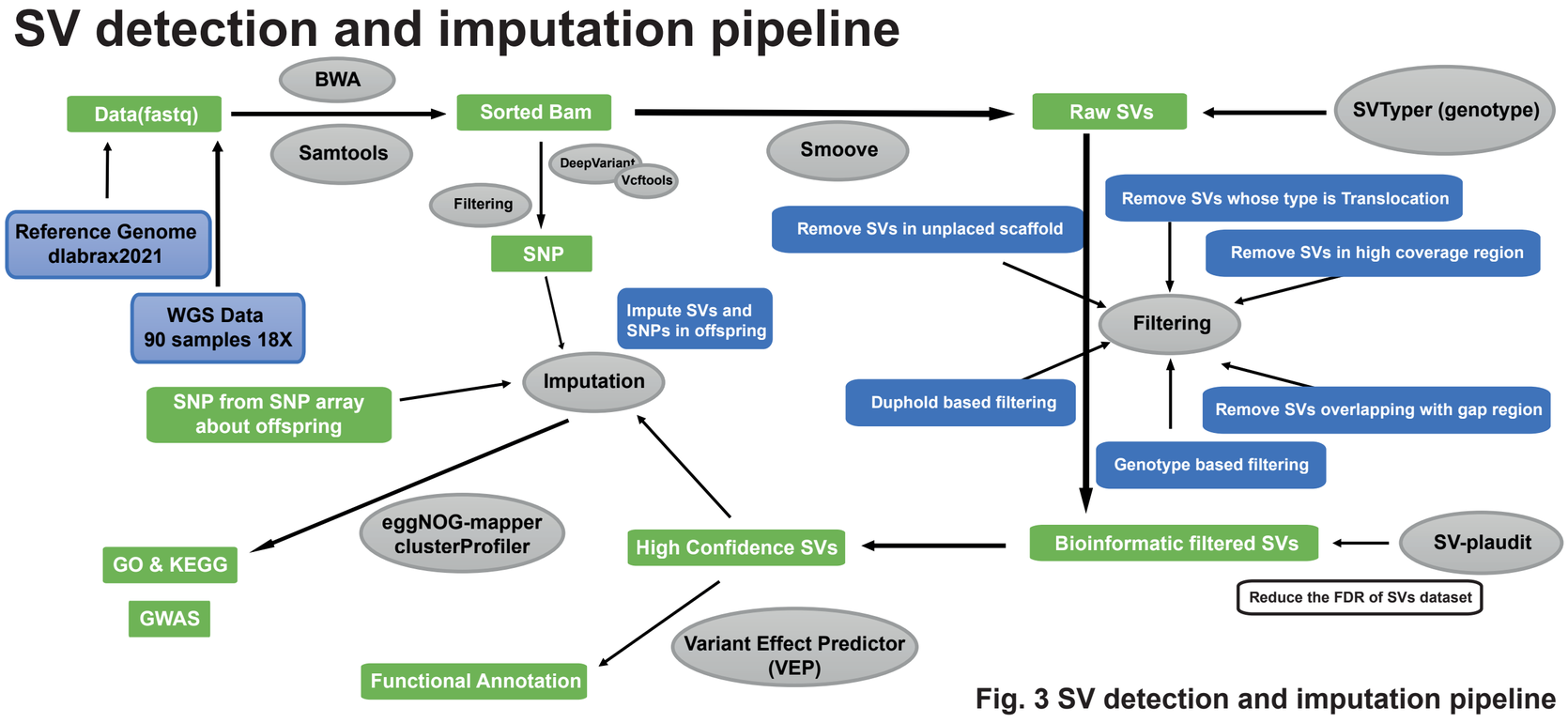

Fig. 3 SV detection and imputation pipeline

The SV landscape of the European Seabass

- **21,428 high-confidence SVs** were identified after rigorous filtering and manual curation, including 21,320 deletions, 75 duplications and 33 inversions (Fig. 4a).
- SV **minor allele frequencies (MAF)** are shown in (Fig 4b) and their **non-uniform distribution** throughout the seabass genome is shown in (Fig 4c). The **length distribution** of SVs is shown in Fig 4d.
- Ensembl Variant Effect Predictor (VEP) was used to annotate the potential impact of SVs potential effect on genes (Fig 5a). Most SVs were found in non-coding regions (Fig. 5b). **492 genes** overlapping SVs were annotated as **coding sequence variants**. After performing GO and KEGG enrichment, the top 10 enriched GO terms and KEGG pathways, ranked by adjusted P-value, are shown in Fig 5c and Fig 5d.

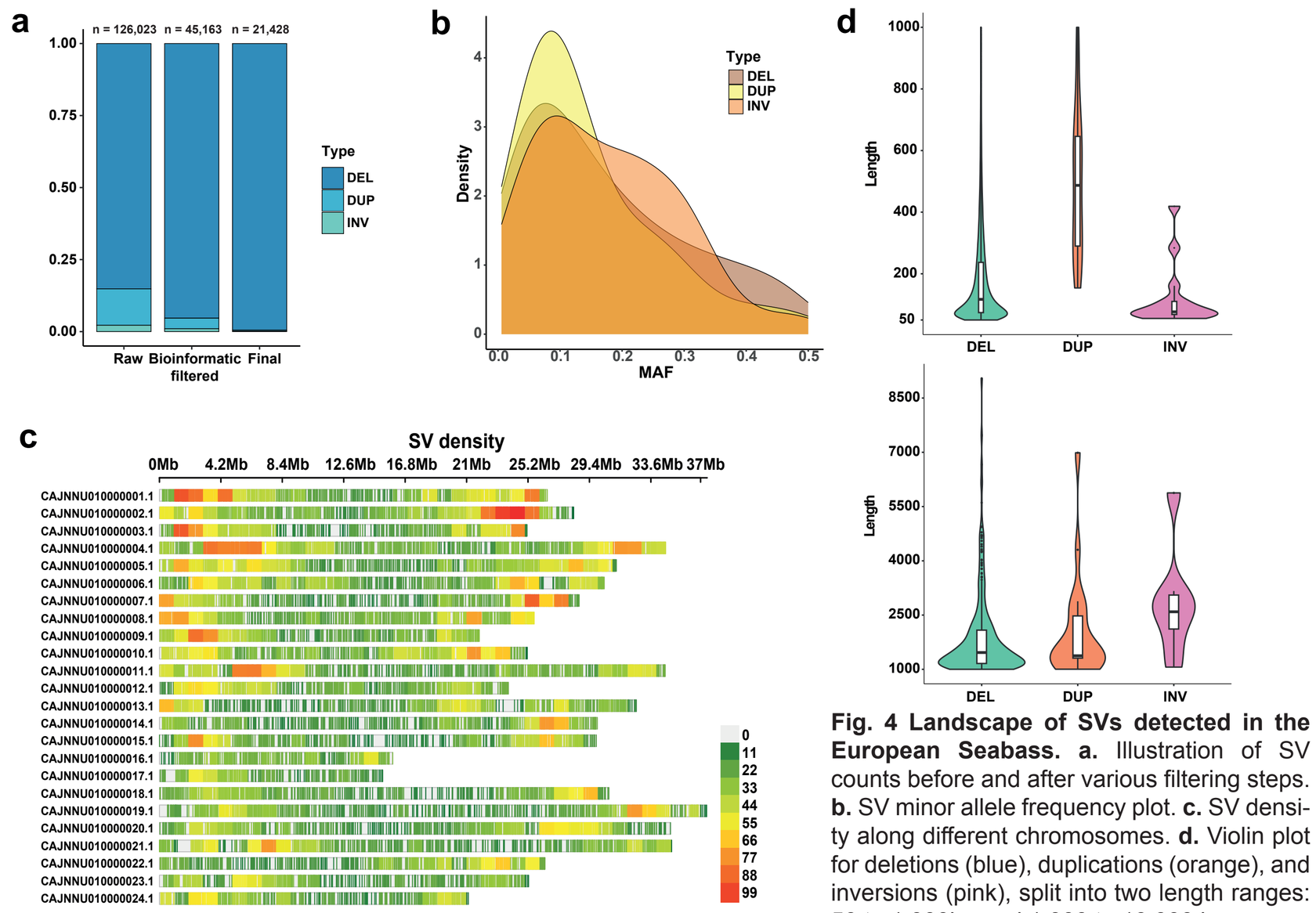

Fig. 4 Landscape of SVs detected in the European Seabass. a. Illustration of SV counts before and after various filtering steps. **b.** SV minor allele frequency plot. **c.** SV density along different chromosomes. **d.** Violin plot for deletions (blue), duplications (orange), and inversions (pink), split into two length ranges: 50 to 1,000bp and 1,000 to 10,000 bp.

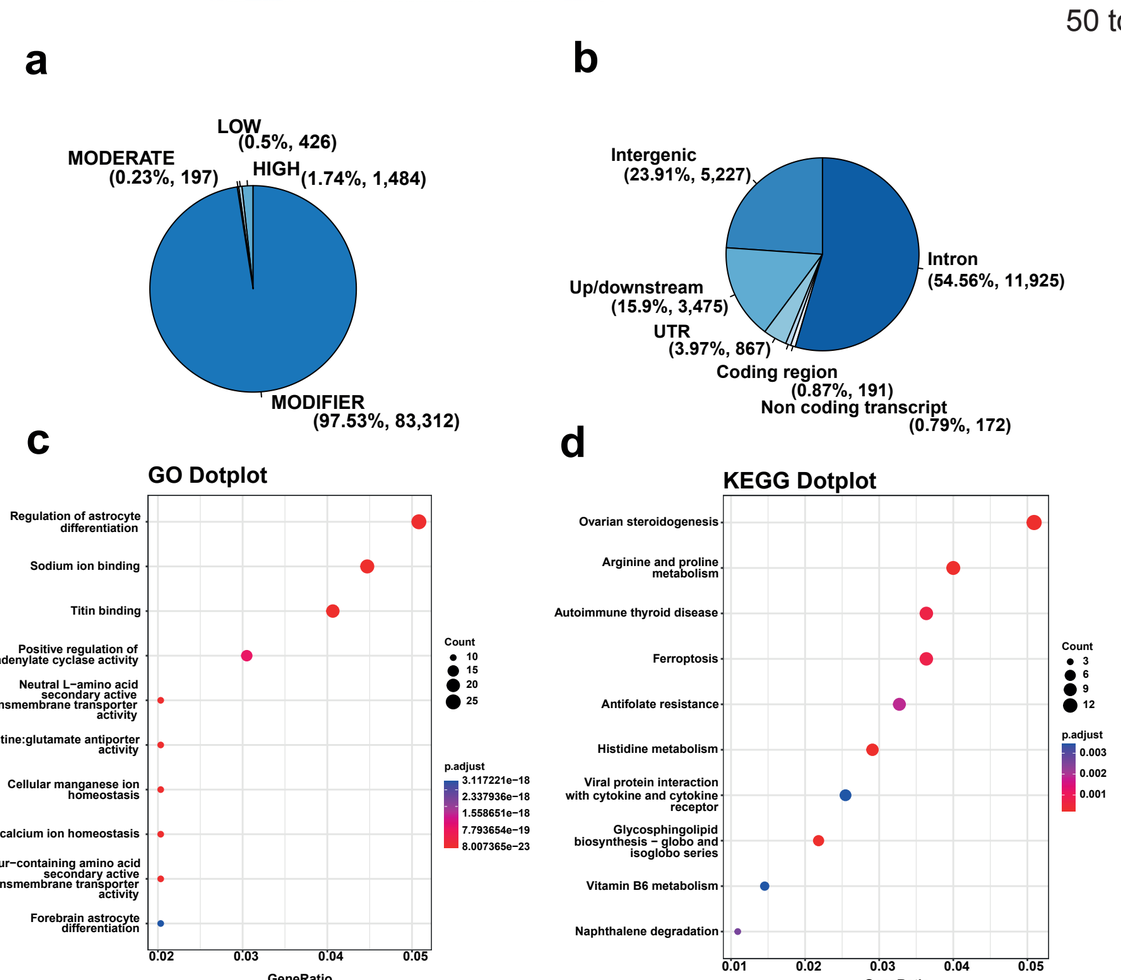

Fig. 5 SV annotation information. a. Pie chart illustrating different SVs effects predicted by SnpEff. **b.** SV annotation according to VEP. **c.** GO enrichment for genes related to the SVs which causing coding regions changes. **d.** KEGG enrichment for genes related to the SVs which causing coding regions changes.

Final perspectives

- Integration with Regulatory Elements: The next step involves integrating identified structural variants (SVs) with **regulatory elements** defined by ATAC-Seq and ChIP-Seq from the AQUA-FAANG project to understand their impact on gene expression.
- Fine Mapping of QTL Regions: Fine **mapping of QTL regions** will help to precisely identify the genetic determinants of economically important traits in European seabass.
- Future Genetics Studies: This study provides a comprehensive SV landscape, serving as a critical resource for future research into the genetic architecture of heritable traits in aquaculture species, aiming to enhance breeding programs and improve aquaculture productivity.

Biotechnology and Biological Sciences Research Council
THE UNIVERSITY of EDINBURGH
AQUA-FAANG

Contact