Statistical Analysis and Description MRR Project

Teddy ALEXANDRE, Arthur BABIN

2022-11-14

Introduction

Over half of the world's population is fed by the cultivars of *Oryza sativa*, also referred to as Asian rice, which is a form of cultivated rice. Hence, rice production and productivity is essential for guaranteeing food security, alleviating hunger, and setting the stage for long-term economic growth.

Furthermore O.Sativa is a model organism for the botany of cereals (small size, ease of growth, high fecundity, short generation time, small genome and amenability to genetic manipulation, including crossing, mutagenesis and gene modification) and has a genome that is *easy to genetically manipulate* with 430 Mbp spread across 12 chromosomes that can be sorted into five sub-populations: - Indica - Aus - Temperate Japonica - Tropical Japonica - Aromatic

The goal is to understand the genetic basis of various physiological, developmental, and morphological traits in order to lay the groundwork for improving rice yield, quality, and sustainability.

genotype dataset:

lignes (36901) : chaque ligne correspond à un SNP colonnes (413+marker+chrom+pos): marker allele data valeur de chaque accession pour le SNP correspondant 0=homozygous (major+major), 1=heterozygous (major+minor), 2=homozygous (minor+minor), NA (represents missing data) - marker: identifiant du SNP (aussi appelé marqueur) - chrom: chromosome sur lequel se situe le SNP - pos: position du SNP dans la séquence ADN du chromosome associé

phenotype dataset

lignes (413) : 1 observation correspondant à une accession colonnes (38) : correspond aux "traits" ie les caractères utilisés pour décrire l'accession (des moyennes sont calculées pour représenter l'accession dans le dataset à partir des différentes plantes de l'accession) - "HybID": numéro d'accession (identifiant unique de l'accession) - "NSFTVID": identifiant local de l'accession (va de 1 à 413) - "Flowering.time.at.Arkansas", "Flowering.time.at.Faridpur", "Flowering.time.at.Aberdeen", "Year07Flowering.time.at.Arkansas", "Year06Flowering.time.at.Arkansas": flowering time of the plant in days - "FT.ratio.of.Arkansas.Aberdeen", "FT.ratio.of.Faridpur.Aberdeen": flowering time ratio

Plant Morphology attributes

- "Culm.habit": estimated average angle of inclination of the base of the main culm (from 0 to 9)
- "Leaf.pubescence": large pubescence (=1) and small pubescence (=0)
- "Flag.leaf.length": length in cm of the flag leaf (the top leaf on a rice stem, closest to the grain-producing panicle)
- "Flag.leaf.width": width in cm of the flag leaf
- "Awn.presence": if the plant has awn (=1) or not (=0)

Yield-related attributes

- "Panicle.number.per.plant": number of panicle per plant
- "Plant.height": height of the plant in cm
- "Panicle.length": length of the panicle in cm
- "Primary.panicle.branch.number": number of panicle branches attached to the main central axis
- "Seed.number.per.panicle": number of seed per panicle
- "Florets.per.panicle": number of florets per panicle
- "Panicle.fertility": fertility ratio of the panicles

Seed morphology attributes

- "Seed.length": seed length in mm
- "Seed.width": seed width in mm
- "Seed.volume": seed volume in dozens of mm³
- "Seed.surface.area": seed surface area in dozens of mm²
- "Brown.rice.seed.length": brown rice (the seed without the husk) length in mm
- "Brown.rice.seed.width": brown rice width in mm
- "Brown.rice.surface.area": brown rice surface area in dozens of mm²
- "Brown.rice.volume": brown rice volume in dozens of mm²
- "Seed.length.width.ratio": ratio between seed length and seed width
- "Brown.rice.length.width.ratio": ratio between brown rice length and brown rice width
- "Seed.color": descriptor of the color of the seed (1 or 0)
- "Pericarp.color": descriptor of the color of the pericarp (1 or 0)

Stress tolerance attributes

- "Straighthead.suseptability": rice susceptibility to straighthead (disease) from 1 (no symptoms) to 9 (severe damages with plants not heading)
- "Blast.resistance": rice resistance to foliar blast (disease) from 0 (Highly resistant) to 9 (Very susceptible)

Quality-related attributes

- "Amylose.content": Amylose content in percentage
- "Alkali.spreading.value": Alkali Spreading Value (ASV) related to the gelatinization temperature
- "Protein.content": Protein content in percentage

Five sub-populations - Indica - Aus - Temperate Japonica - Tropical Japonica - Aromatic