

XV JORNADAS DE
BIOLOGIA DE LEVEDURAS
"PROFESSOR NICOLAU VAN UDEN"

Porto – 15/16.06.2007

BIOINFORMATIC APPROACHES FOR THE GENETIC AND PHENOTYPIC CHARACTERIZATION OF A *SACCHAROMYCES CEREVISIAE* WINE YEAST COLLECTION

Ricardo Franco-Duarte, Margarida Casal, Dorit Schuller

Departamento/Centro de Biologia, Universidade do Minho



Introduction

- ↑ *Saccharomyces cerevisiae* is the most well known eukaryotic model, mankind's oldest domesticated organism and the world's premier commercial microorganism for biotechnological applications
- ↑ Genetic/genomic studies are traditionally carried out with a few laboratory strains and natural isolates are still poorly investigated
- ↑ With the advent of molecular typing methods, genetic variability and relatedness among natural isolates of *S. cerevisiae* can be established, and linked to phenotypic characteristics
- ↑ Bridging the genotype-to-phenotype gap using bioinformatics approaches will become very important for the industrial scale production for biotechnologically valuable products

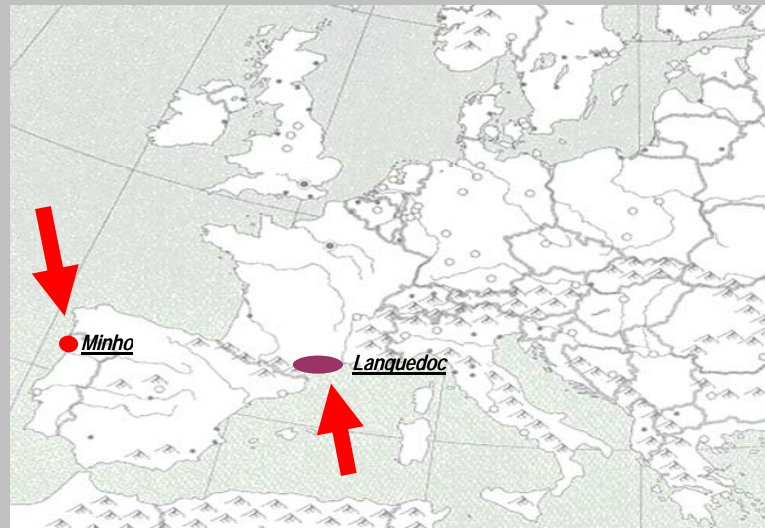


Objectives

- 🕒 To compose and systematize a *Saccharomyces cerevisiae* strain collection including genetic profiles from multiple molecular typing methods and phenotypic information
- 🕒 To investigate genetic relatedness among natural populations
- 🕒 To compare genetic and phenotypic variation using bioinformatics approaches as a basis for future analysis to estimate a strain's biotechnological potential



Biogeographical large-scale survey – *Saccharomyces cerevisiae* Wine Yeast Collection

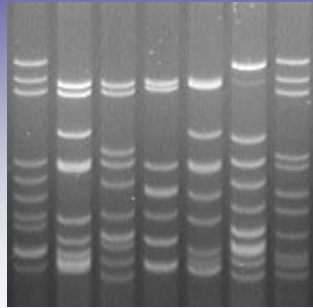


	<i>Minho</i> 2001 - 2006	<i>Languedoc</i> 2001-2003	<i>Total</i>
Number of isolates	2520	2160	6380
<i>S. cerevisiae</i> strains (unique genetic profiles)	340	100	440



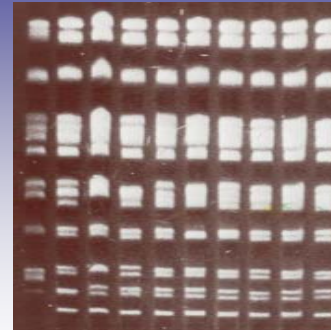
Molecular methods for *S. cerevisiae* strain characterization

Mitochondrial DNA restriction analysis (mtDNA RFLP)



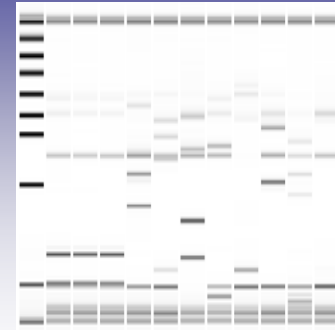
(Quero et al., 1992;
Lopez et al., 2001)

Electrophoretic karyotyping (PFGE)



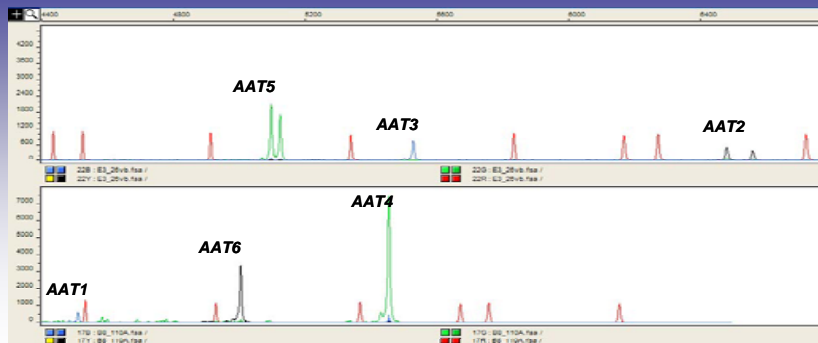
(Carle and Olson., 1995;
Blondin and Vezinhet, 1988)

PCR based -Interdelta sequence amplification



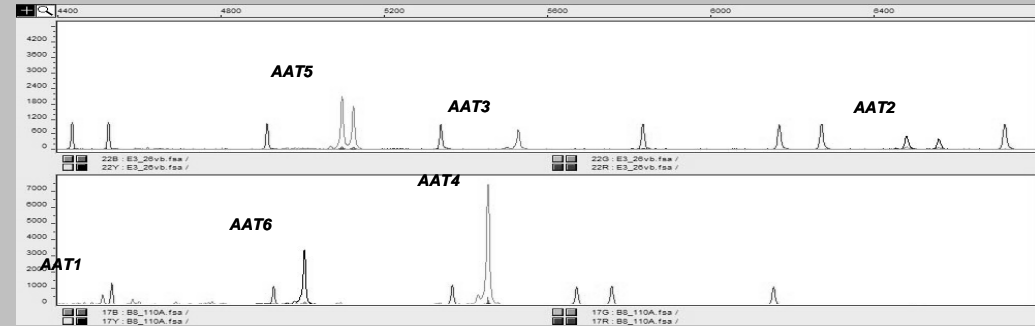
(Ness et al., 1993)
(Legras et al., 2003)

Microsatellite analysis



Molecular methods for *S. cerevisiae* strain characterization

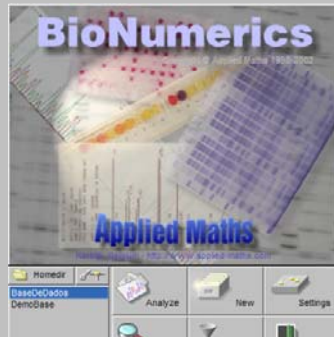
Microsatellite analysis



Microsatellite	Chromosome	Position/ORF	Repeat	N° of alleles	Ref.
ScAAT1	XIII	86 901 – 87 129	ATT	29	Pérez and Gallego, 2001
ScAAT2	II	YBL084c	ATT	18	
ScAAT3	IV	YDR160w	ATT	19	
ScAAT4	VII	431 334 – 431 637	ATT	17	
ScAAT5	XVI	897 028 - 897 259	TAA	6	
ScAAT6	IX	105 661 – 105 926	TAA	10	
YPL009	XV	YOR156c	TAA	13	Legras et al., 2005
SCYOR267C	XV	YOR267c	TGT	12	
C4	XV	110 701-110 935	TAA+TAG	9	
C5	VI	210250-210414	GT	19	
C11	X	518870-519072	GT	18	



The BIONUMERICs software for databasing and cluster analysis



BioNumerics

File Edit Database Subsets Experiments Comparison Identification Scripts

Complete database

Key	Ano	Quinta	a/d	Ano 2	Quinta 2	a/d 2	Ano 3	Quinta 3	a/d 3	Ano 4	Quinta 4	1	2	3	4	5	6	
1	Estirpe 001	2001	Port C	a	2001	Port C	d	2003	Port C	d								
2	Estirpe 002	2001	Port C	a														
3	Estirpe 008	2001	Port C	a	2001	Port C	d											
4	Estirpe 009	2001	Port C	d														
5	Estirpe 010	2001	Port C	a	2001	Port C	d	2001	Port P	d	2002	Port P						
6	Estirpe 013	2001	Port C	a	2001	Port C	d	2002	Port P	d	2003	Port C						
7	Estirpe 018	2001	Port C	a	2001	Port C	d	2001	Port P	d								
8	Estirpe 025	2001	Port C	a														

No.	ENZYME ASSAYED FOR	SUBSTRATE	pH	RESULT	
				POSITIVE	NEGATIVE
1	Control			Colorless or color of the sample if it has an intense coloration	
2	Alkaline phosphatase	2-naphthyl phosphate	8.5	Violet	
3	Esterase (C 4)	2-naphthyl butyrate	6.5	Violet	
4	Esterase Lipase (C 8)	2-naphthyl caprylate	7.5	Violet	
5	Lipase (C 14)	2-naphthyl myristate	"	Violet	
6	Leucine arylamidase	L-leucyl-2-naphthylamide	"	Orange	
7	Valine arylamidase	L-valyl-2-naphthylamide	"	Orange	
8	Cystine arylamidase	L-cystyl-2-naphthylamide	"	Orange	
9	Trypsin	N-benzoyl-DL-arginine-2-naphthylamide	8.5	Orange	
10	α -chymotrypsin	N-glutaryl-phenylalanine-2-naphthylamide	7.5	Orange	
11	Acid phosphatase	2-naphthyl phosphate	5.4	Violet	Colorless or Very pale yellow *
12	Naphthol-AS-BI-phosphohydrolase	Naphthol-AS-BI-phosphate	"	Blue	
13	α -galactosidase	6-Br-2-naphthyl- α D-galactopyranoside	"	Violet	
14	β -galactosidase	2-naphthyl- β D-galactopyranoside	"	Violet	
15	β -glucuronidase	Naphthol-AS-BI- β D-glucuronide	"	Blue	
16	α -glucosidase	2-naphthyl- α D-glucopyranoside	"	Violet	
17	β -glucosidase	6-Br-2-naphthyl- β D-glucopyranoside	"	Violet	
18	N-acetyl- β -glucosaminidase	1-naphthyl-N-acetyl- β D-glucosaminide	"	Brown	
19	α -mannosidase	6-Br-2-naphthyl- α D-mannopyranoside	"	Violet	
20	α -fucosidase	2-naphthyl- α L-fucopyranoside	"	Violet	

Experiments

Files

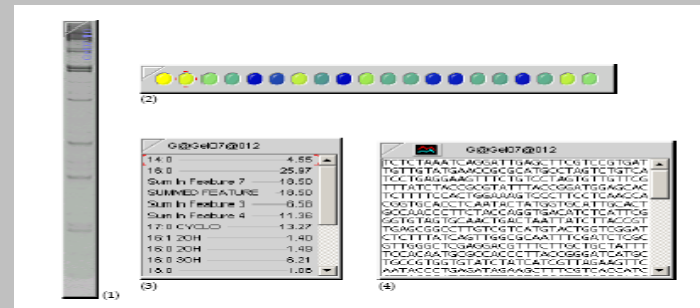
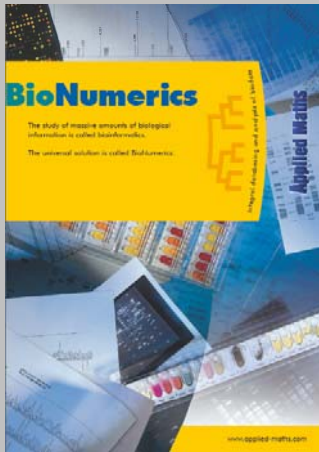
A_1
A_2
B_1
B_2
BioCant
C_1
C_2
D_1
D_2

Libraries

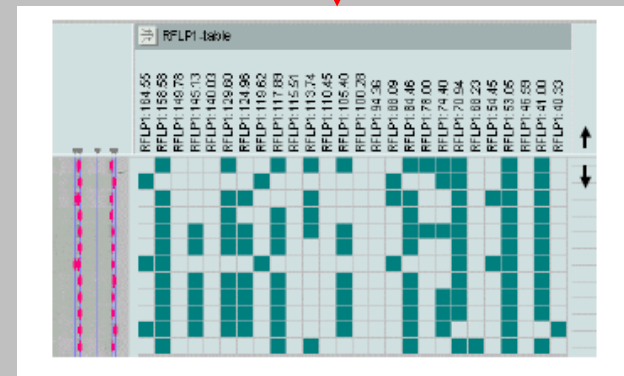
clust)
stras)
stras)
mstras)



The BIONUMERICS software for databasing and cluster analysis

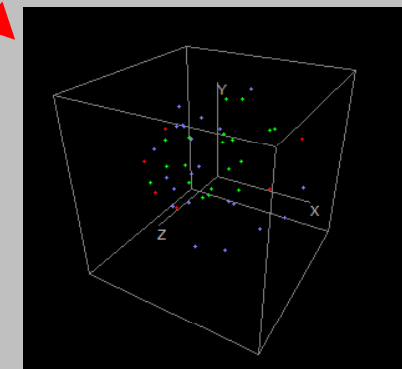
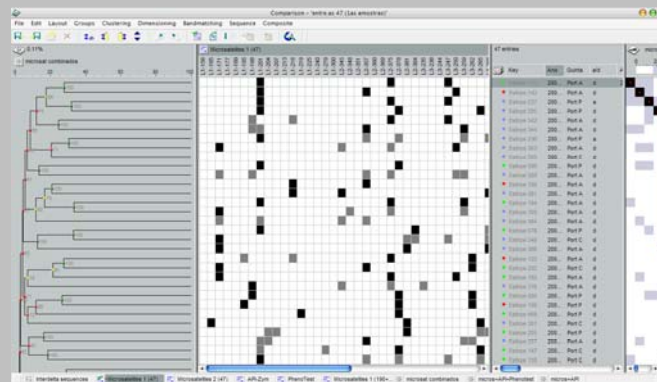


Entries from multiple fingerprint and character types



Binary/numerical value experiment file(s)

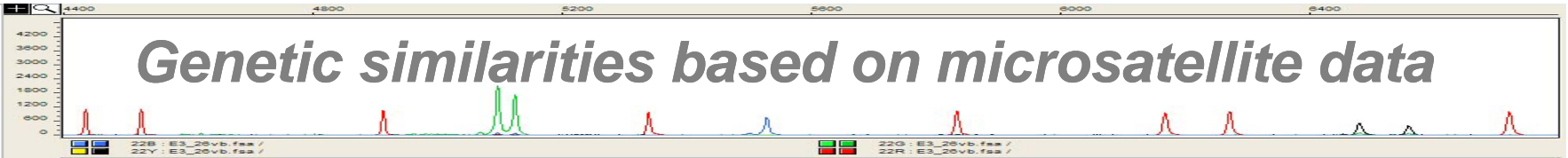
Approaches with single/multiple (polyphasic) fingerprint and character types



XV JORNADAS DE BIOLÓGIA DE LEVEDURAS "PROFESSOR NICOLAU VAN UDEN"

Porto, 15/16 Junho 2007

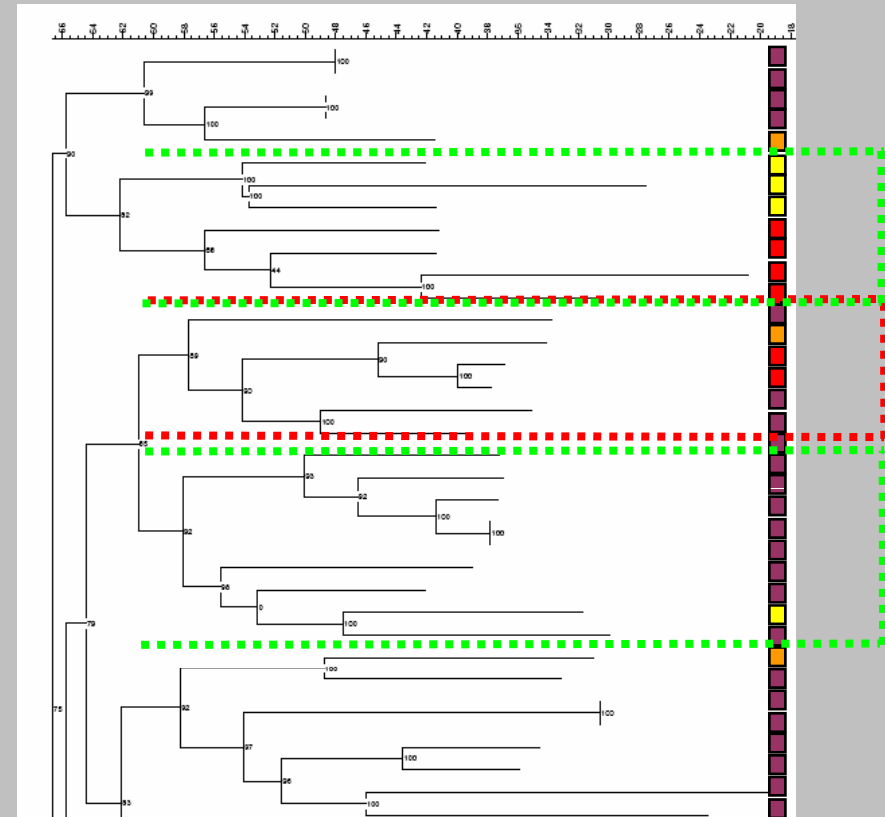
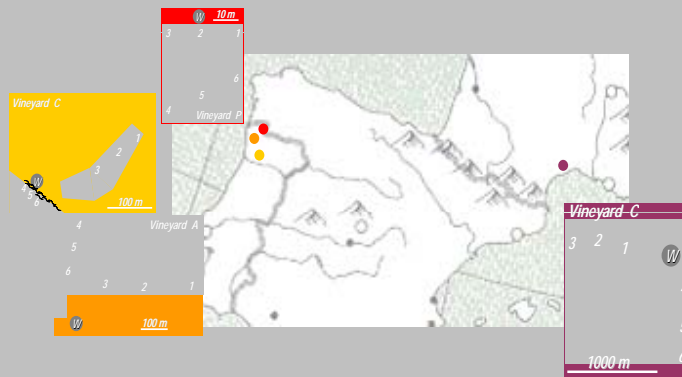
Genetic similarities based on microsatellite data



Analysis of individual strains

What are the genetic differences among strains from distinct geographical regions?

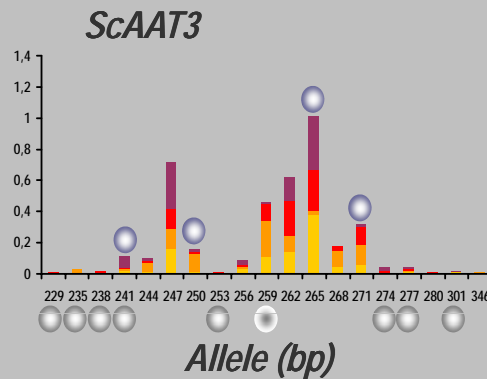
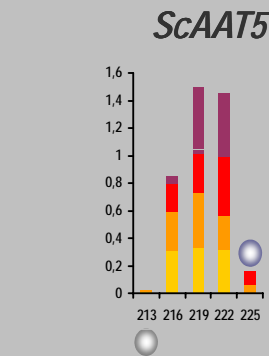
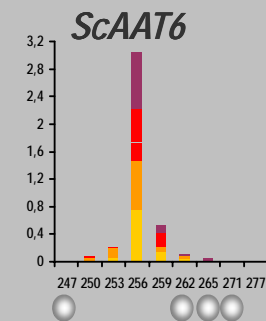
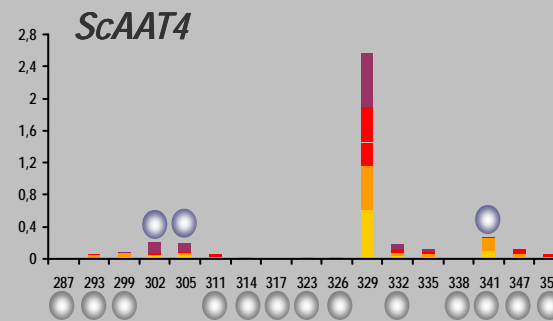
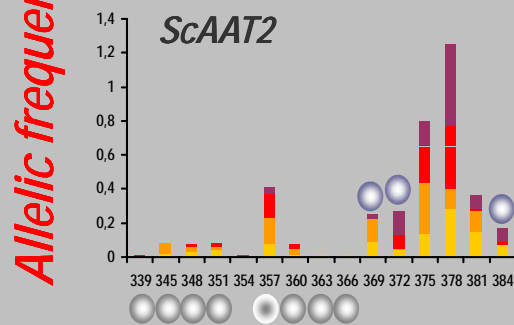
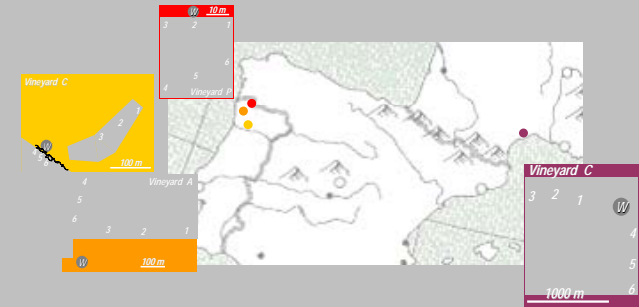
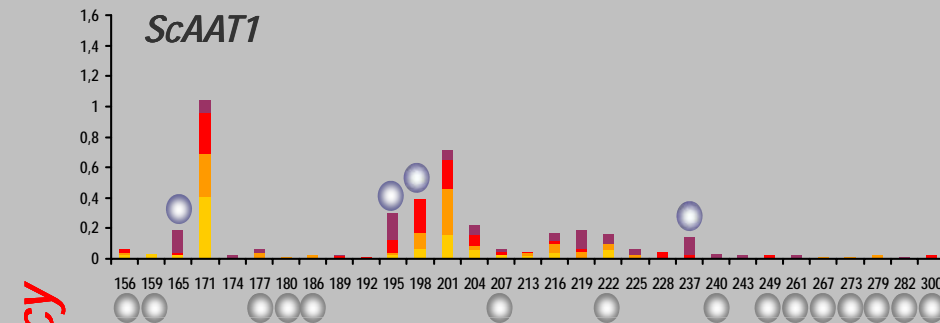
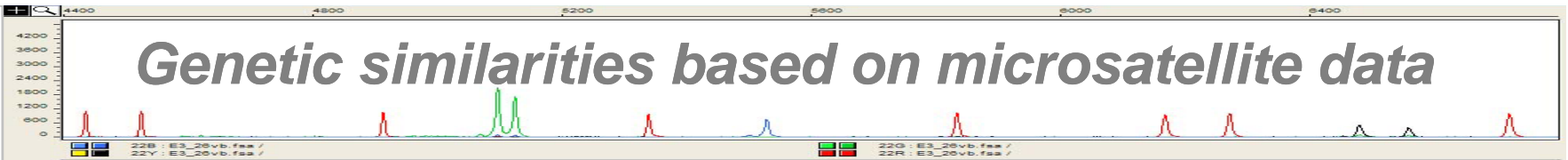
Correlation between geographic distance and genetic similarity?






Neighbor-joining clustering
 $r=0,75$



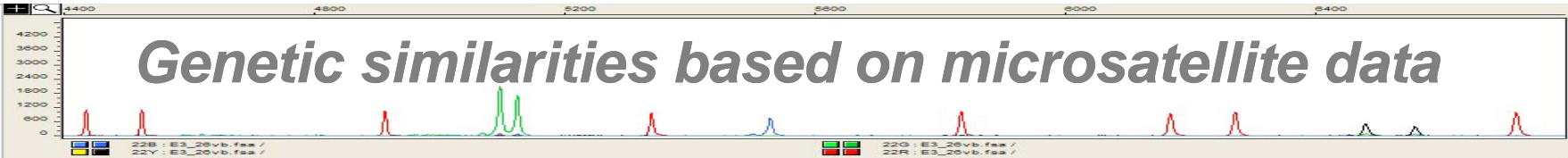
Genetic similarities based on microsatellite data



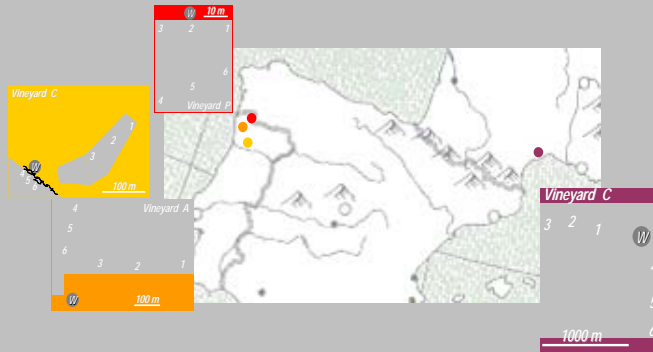
-  *New allele*
-  *Vineyard – specific allele*
-  *Wine Region – specific allele*



Genetic similarities based on microsatellite data



Comparison of populations



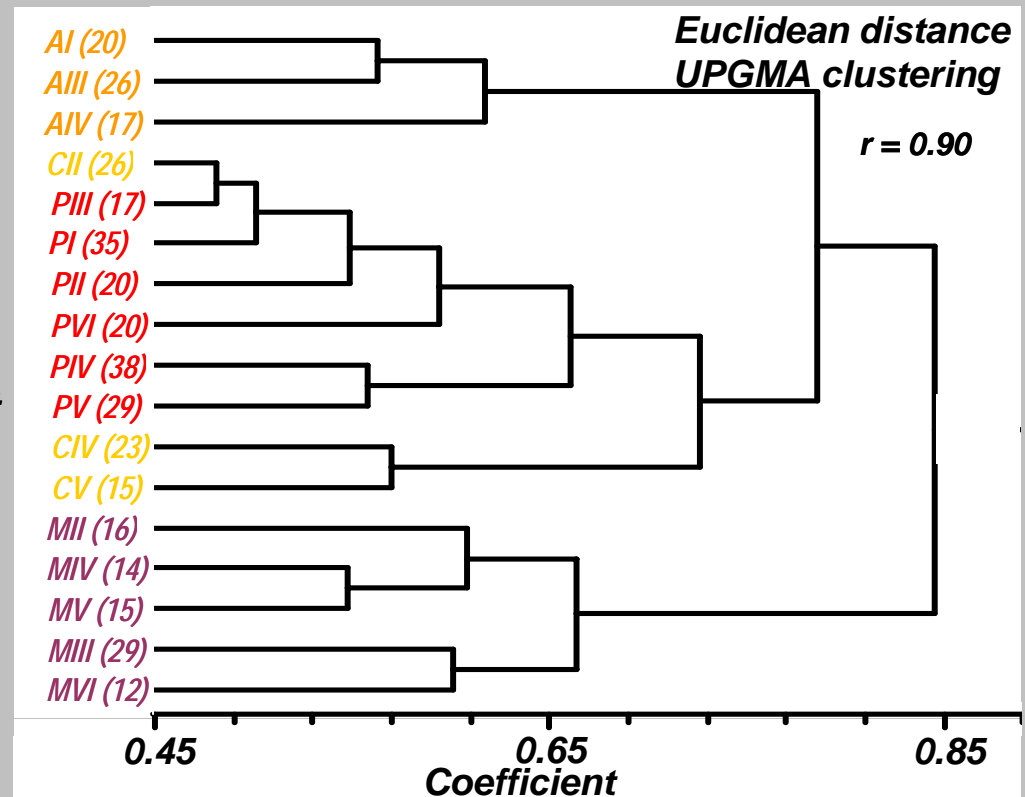
Populational substructure



Genetic differences delimit specific populations - characteristic for each vineyard



Correlation between geographic distance and genetic similarity ?



() number of strains analyzed

Schuller and Casal, 2007



Comparison between genetic and phenotypic variability

Neural networks (SOM)

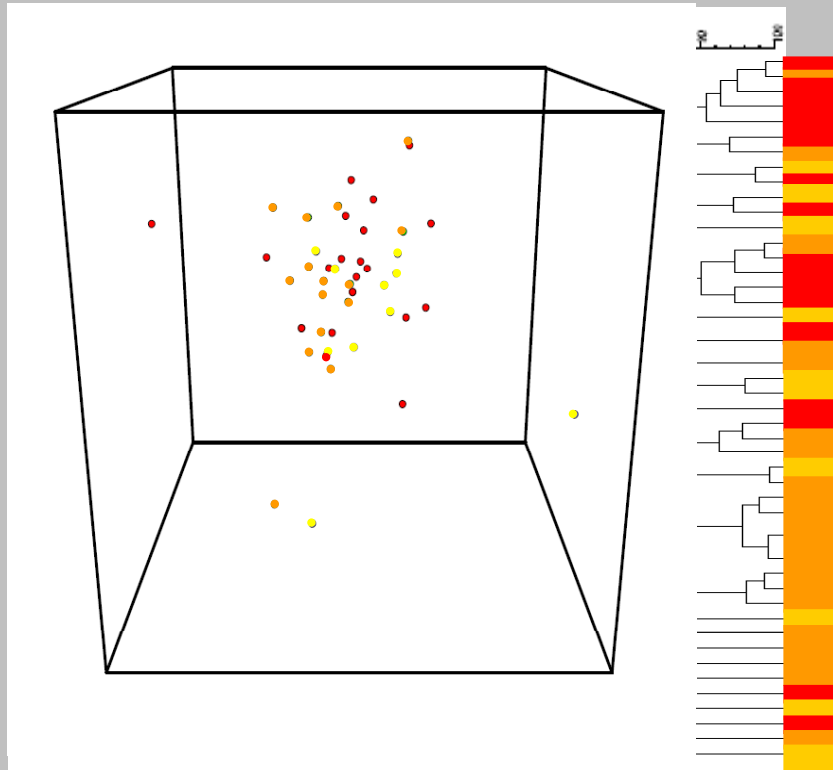
Epoch 25 Clustering factor: 9.0

Selection of a genetically most diverse set of strains



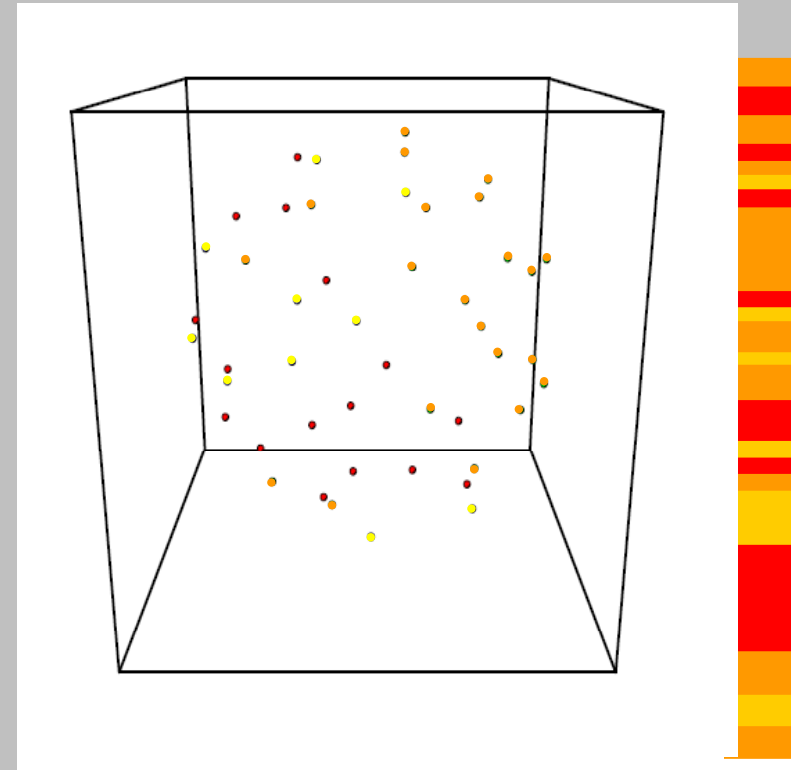
Comparison between genetic and phenotypic variability

API-ZYM analysis



Pearson correlation
UPGMA clustering
 $R=0,80$

Microsatellite analysis



Pearson correlation
UPGMA clustering
 $R=0,64$



Ongoing work and future perspectives

||- Include additional phenotypic data with biotechnological relevance (for example wine making), using most objective phenotyping approaches

||- Expand computational methods (artificial intelligence) in order to make predictions regarding biotechnological relevant characteristics of natural strains



Acknowledgements

**Universidade do Minho
Biology Department**

Dorit Schuller

Margarida Casal

**INRA Montpellier
Sciences pour
Oenologie**

Sylvie Dequin

Brigitte Cambon

**Department of
Química University
Nova of Lisbon**

João Aires de Sousa

Financially Supported by the POCI 2010 program (Feder/FCT, POCI/AGR/56102/2004)

 **Ciência.Inovação
2010** Programa Operacional Ciência e Inovação 2010
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR



UNIAO EUROPEIA
FUNDOS ESTRUTURAIS



XV JORNADAS DE
BIOLOGIA DE LEVEDURAS
"PROFESSOR NICOLAU VAN UDEN"

Porto, 15/16 Junho 2007