Genome-wide characterization of gene-families in Resequencing projects – *Challenges and Perspectives*

Maharajah Ponnaiah

E P G V

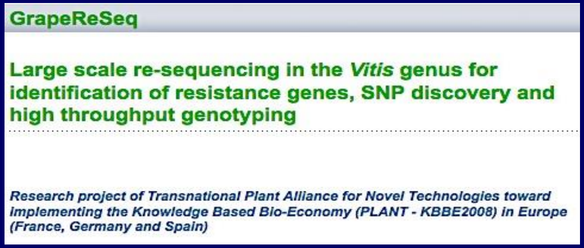
Etude du **P**olymorphisme des **G**énomes **V**égétaux

US1279 - BAP

Introduction

- *Genomics Era* – Low Cost - Resequencing large genotypes
- More NGS tools for data mining – Mapping – SNP detection – downstream analysis.
- Paucity of robust tools in the context of gene families and distantly placed species from the reference genome.
- Gene family size variation is an important mechanism that shapes the natural variation for adaptation in various species.
- The pattern of gene family size variation in plant species are still not well understood.

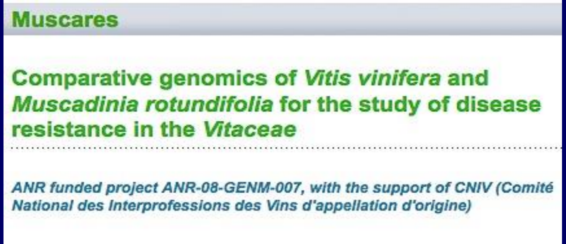
Vitis and Muscadine species



GrapeReSeq

Large scale re-sequencing in the *Vitis* genus for identification of resistance genes, SNP discovery and high throughput genotyping

Research project of Transnational Plant Alliance for Novel Technologies toward implementing the Knowledge Based Bio-Economy (PLANT - KBBE2008) in Europe (France, Germany and Spain)



Muscares

Comparative genomics of *Vitis vinifera* and *Muscadinia rotundifolia* for the study of disease resistance in the *Vitaceae*

ANR funded project ANR-08-GENM-007, with the support of CNIV (Comité National des Interprofessions des Vins d'appellation d'origine)

Vitis vinifera	*Vitis sylvestris* (wild)	*Vitis* (other) *species*	*Muscadinia species*
Araklinos	*Dirmstein-male*	*V.aestivalis.aestivalis-1*	*M.carlos*
Carbernet-franc	*Lambrusque-campmarcel*	*V.aestivalis.aestivalis-2*	*M.Dulcet*
Carignan-noir	*Lambrusque-PSL2*	*V.aestivalis.aestivalis-3*	*M.fry*
Castellana-blanca	*Teulere-pied-sauvage*	*V.amurensis.colmar*	*M.mchx*
Chardhchi-blanc		*V.amurensis.rupprecht*	*Regale*
Chouchillon		*V.berlandieri.planchon*	*M.trayshed*
Colorino		*V.berlandieri-Resseguier*	*M.yugax-carlos*
Espadeiro-tinto		*V.berlandieri-10594*	
Hebén		*V.cinerea.cinerea-1*	
Jaén		*V.cinerea.cinerea-2*	
Lambrusque-E		*V.cinerea.cinerea-3*	
Listàn-Prieto		*V.labrusca.concord*	
Malvasia-di-Sardegna		*V.labrusca.fredonia*	
Maska		*V.lincecumii*	
Médouar			
Orlovi-nogti			
PN40024 -TruSeq			
Riesling			
Savagnin			
Tsolikoouri			
20	**4**	**14**	**7**
Total	**45**		

Mapping Assembly Tool

CLC GENOMICS WORKBENCH

A comprehensive and user-friendly analysis package for analyzing, comparing, and visualizing next generation sequencing data.

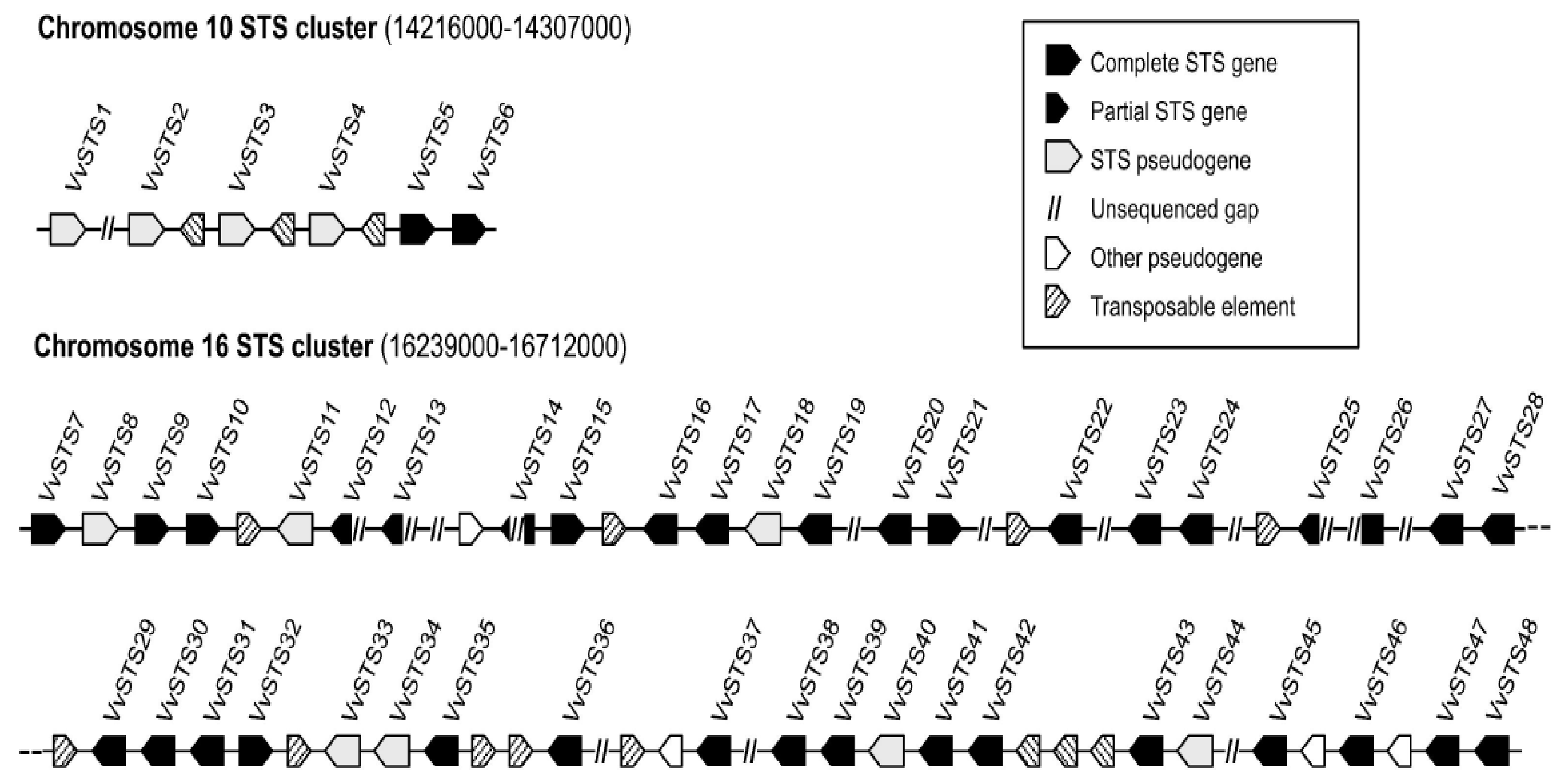
CLC ASSEMBLY CELL

Fast and accurate command-line tool for read mapping or *de novo* assembly of large and complex NGS datasets.

Stilbene Synthase (STS)

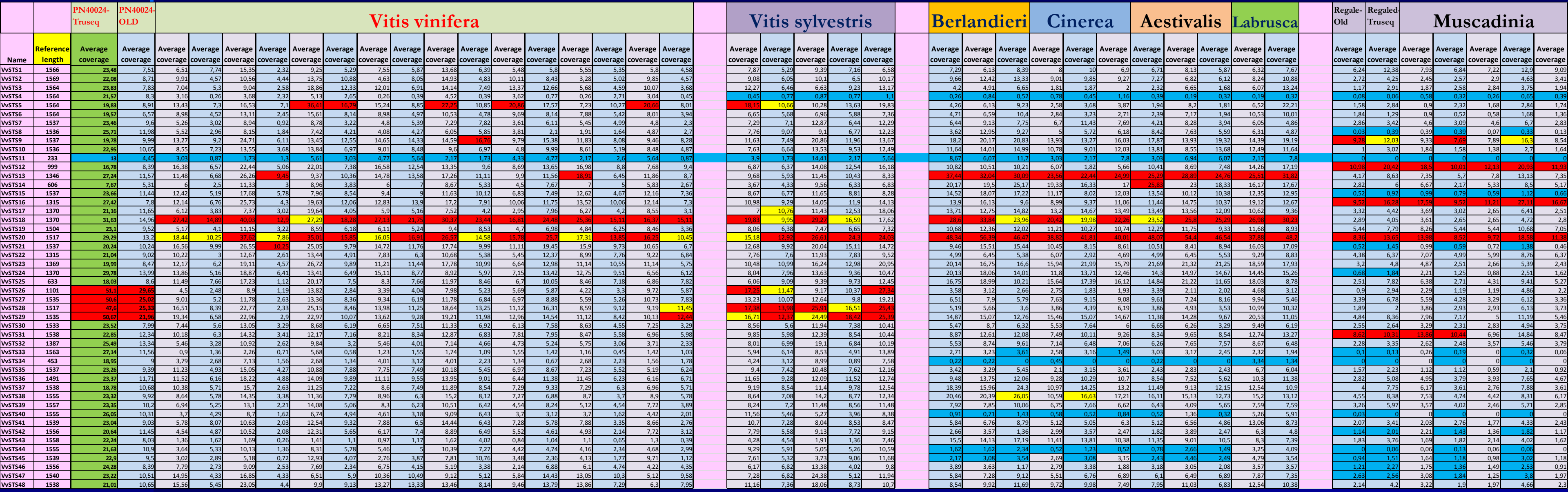
- ☐ Stilbenes belong to a small family of phenlypropanoids
- ☐ Produced in numerous plant sps. against several biotic and abiotic stresses, including to UV and chemical treatments.
- ☐ Stilbenes such as resveratrol represent an imperative medicinal property –
  - Anticancer
  - Antinflamation
  - Diabetes
  - Anti-Ageing (*French Paradox*)
- ☐ Mysterically grapevine possess numerous STS genes (~ 45 genes) unlike in other plant species, makes the molecular evolutionist and plant breeders ever more curious.

The STS clusters were annotated in the reference grape genome PN40024



URGV-INRA Adopted from – Parage *et al* (2012) Plant Physiology

CLC Mapping Statistics



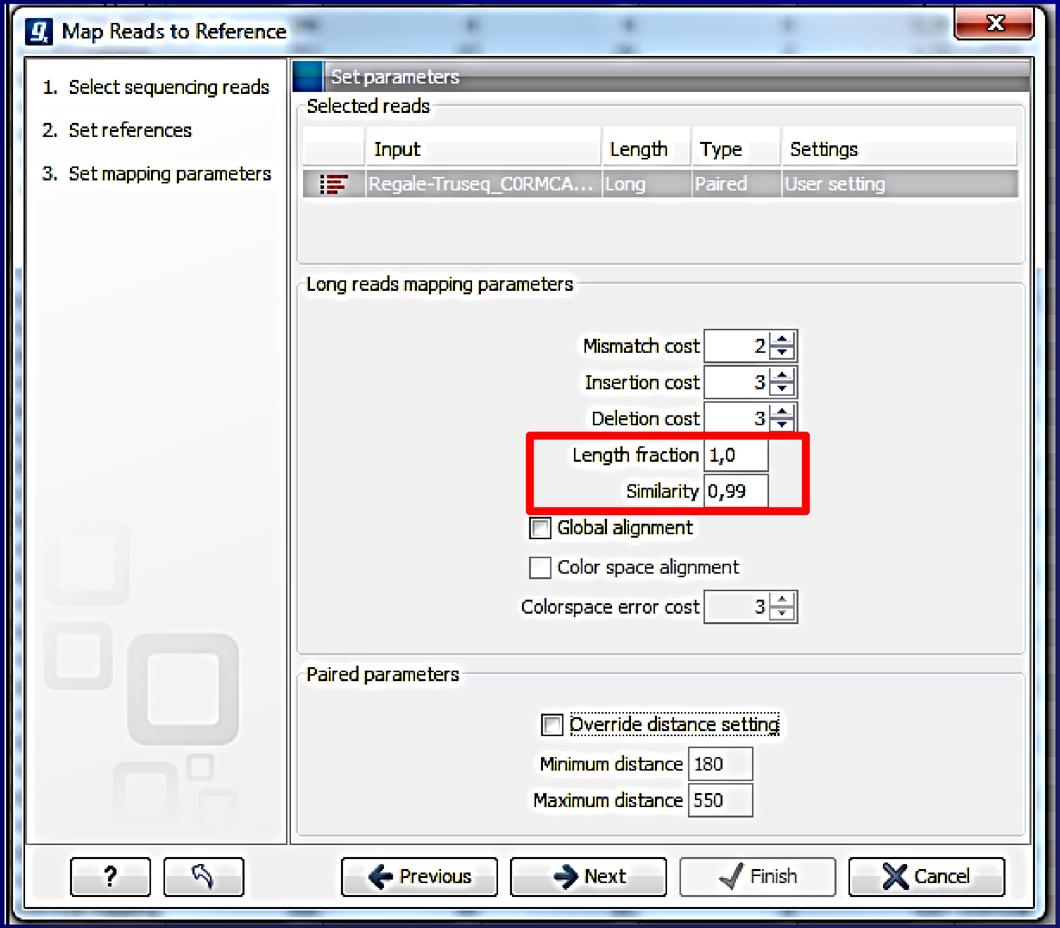
		PN40024-Truseq	PN40024-OLD	Vitis vinifera																	Vitis sylvestris						Berlandieri			Cinerea			Aestivalis			Labrusca			Regale-Old	Regaled-Truseq	Muscadinia				
Name	Reference length	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage		Average coverage	Average coverage	Average coverage	Average coverage	Average coverage		Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage		Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage	Average coverage
VvSTS1	1566	23,48	7,51	6,51	7,74	15,35	2,32	9,25	5,29	7,55	5,87	13,68	6,39	5,48	5,8	5,55	5,35	5,8	4,58		7,87	5,29	9,39	7,16	6,58		7,29	6,13	8,39	8	10	6,9	6,71	8,13	5,87	6,32	7,67		6,24	12,38	7,93	6,84	7,22	12,9	9,09
VvSTS2	1569	22,08	8,71	9,91	4,57	10,56	4,44	13,75	10,88	4,63	8,05	14,93	4,83	10,11	8,43	3,28	5,02	9,85	4,57		9,08	6,05	10,1	6,5	10,17		9,66	12,42	13,33	9,01	9,85	9,27	7,27	6,82	6,12	8,24	10,88		2,72	4,25	2,45	2,57	2,9	4,63	3,41
VvSTS3	1564	23,83	7,83	7,04	5,3	9,04	2,58	18,86	12,33	12,01	6,91	14,14	7,49	13,37	12,66	5,68	4,59	10,07	3,68		12,27	6,46	6,63	9,23	13,17		4,2	4,91	6,65	1,81	1,87	2	2,32	6,65	1,68	6,07	13,24		1,17	2,91	1,87	2,58	2,84	3,75	1,94
VvSTS4	1564	21,57	8,3	3,16	0,26	3,68	2,32	5,13	2,65	0,26	0,39	4,52	0,39	3,62	0,77	0,26	2,71	3,04	0,45		0,45	0,77	0,87	0,77	1,1		0,26	0,84	0,52	0,78	0,45	1,16	0,39	0,19	0,32	0,19	0,32		0,08	0,06	0,58	0,32	0,26	0,65	0,39
VvSTS5	1564	19,83	8,91	13,43	7,3	16,53	7,1	36,41	16,79	15,24	8,85	27,25	10,85	20,86	17,57	7,23	10,27	20,66	8,01		18,15	10,66	10,28	13,63	19,83		4,26	6,13	9,23	2,58	3,68	3,87	1,94	8,2	1,81	6,52	22,21		1,58	2,84	0,9	2,32	1,68	2,84	1,74
VvSTS6	1564	19,57	6,57	8,98	4,52	13,11	2,45	15,61	8,14	8,98	4,97	10,53	4,78	9,69	8,14	7,88	5,42	8,01	3,94		6,65	5,68	6,96	5,88	7,36		4,71	6,59	10,4	2,84	3,23	2,71	2,39	7,17	1,94	10,53	10,01		1,84	1,29	0,9	0,52	0,58	1,68	1,36
VvSTS7	1537	23,46	9,6	5,26	3,02	8,94	0,92	8,78	3,22	4,8	5,39	7,29	7,82	3,61	6,11	5,45	4,99	4,8	2,3		7,29	7,1	12,87	6,44	12,29		6,44	9,13	7,75	6,7	11,43	7,69	4,21	8,28	3,94	6,05	4,86		2,86	3,42	4,6	3,09	4,6	6,7	2,83
VvSTS8	1536	25,71	11,98	5,52	2,96	8,15	1,84	7,42	4,21	4,08	4,27	6,05	5,85	3,81	2,1	1,91	1,64	4,87	2,7		7,76	9,07	9,1	6,77	12,23		3,62	12,95	9,27	5	5,72	6,18	8,42	7,63	5,59	6,31	4,87		0,03	0,39	0,39	0,39	0,07	0,33	0,13
VvSTS9	1537	19,78	9,99	13,27	9,2	24,71	6,11	13,45	12,55	14,65	14,33	14,59	16,76	9,79	15,38	11,83	8,08	9,46	8,28		11,63	7,49	20,86	11,96	13,67		18,2	20,17	20,83	13,93	13,27	16,03	17,87	13,93	19,32	14,39	19,19		9,28	12,03	9,33	7,69	7,89	16,3	8,54
VvSTS10	1536	22,95	10,65	8,55	7,23	13,55	3,68	13,84	6,97	9,01	8,48	9,6	6,97	4,8	9,99	8,61	5,19	8,48	4,87		7,63	6,64	13,53	9,53	12,49		11,64	14,01	14,99	10,78	9,01	12,03	13,81	8,55	13,68	12,49	11,64		1	3,02	1,84	1,58	1,38	2,7	1,64
VvSTS11	233	13	4,45	3,03	0,87	1,73	1,3	5,61	3,03	4,77	5,64	2,17	1,73	4,33	4,77	2,17	2,6	5,64	0,87		3,9	1,73	14,41	2,17	5,64		8,67	6,07	11,7	3,03	2,17	7,8	3,03	6,94	6,07	2,17	7,8		0	0	0	0	0	0	0
VvSTS12	999	16,78	8,39	16,38	6,57	22,44	5,06	22,01	7,38	16,58	12,54	13,35	9,6	8,69	13,65	16,98	8,8	7,68	9,4		6,87	6,37	14,08	12,54	16,18		10,82	10,51	10,21	6,07	1,82	5,66	10,41	8,69	7,48	14,26	17,19		10,98	20,42	18,5	10,01	12,13	20,93	11,93
VvSTS13	1346	27,24	11,57	11,48	6,68	26,26	9,45	9,37	10,36	14,78	13,58	17,26	11,11	9,9	11,56	18,91	6,45	11,86	8,7		9,68	5,93	11,45	10,43	8,33		37,44	32,04	30,09	23,56	22,44	24,99	25,29	28,89	24,76	25,51	31,82		4,17	8,63	7,35	5,7	7,8	13,13	7,35
VvSTS14	606	7,67	5,31	6	2,5	11,33	3	8,96	3,83	6	7	8,67	5,33	4,5	7,67	7	5	5,83	2,67		3,67	4,33	9,56	6,33	6,83		20,17	19,5	25,17	19,33	16,33	17	25,83	23	18,33	16,17	17,67		2,82	6	6,67	2,17	5,33	8,5	5,17
VvSTS15	1537	23,66	11,44	12,42	5,19	17,68	5,78	7,96	8,54	9,4	9	11,63	10,12	6,83	7,49	12,62	4,67	12,16	7,36		8,67	6,77	11,65	8,81	8,28		14,52	18,07	17,22	11,17	8,02	12,03	13,54	10,12	10,38	12,35	12,95		0,52	0,92	0,99	0,79	0,59	1,12	0,66
VvSTS16	1315	27,42	7,8	12,14	6,76	25,73	4,3	19,63	12,06	12,83	13,9	17,2	7,91	10,06	11,75	13,52	10,06	12,14	7,3		10,98	9,29	14,05	11,9	14,13		13,9	16,13	9,6	8,99	9,37	11,06	11,44	14,75	10,37	19,12	12,67		9,52	16,28	17,59	9,52	11,21	27,11	16,67
VvSTS17	1370	21,16	11,65	6,12	3,83	7,37	3,02	19,64	4,05	5,9	5,16	7,52	4,2	2,95	7,96	6,27	4,2	8,55	3,1		7	10,76	11,43	12,53	18,06		13,71	12,75	14,82	13,2	14,67	13,49	13,49	13,56	12,09	10,62	9,36		3,32	4,42	3,69	3,02	2,65	6,41	2,51
VvSTS18	1370	31,63	14,96	27,42	14,89	40,03	12,9	27,29	18,28	27,13	21,75	30,37	23,44	16,81	24,48	25,36	15,11	16,37	15,11		19,83	9,95	29,27	16,59	17,62		28,6	33,84	23,96	20,42	19,98	22,26	23,52	25,8	25,29	26,98	30,23		2,89	4,05	3,61	2,65	2,65	4,72	2,8
VvSTS19	1504	23,1	9,52	5,17	4,1	11,15	3,22	8,59	6,18	6,11	5,24	9,4	8,53	4,7	6,98	4,84	6,25	8,46	3,36		8,06	6,38	7,47	6,65	7,32		10,68	12,36	12,02	11,21	10,27	10,74	12,29	11,75	9,33	11,68	8,93		5,44	7,79	8,26	5,44	5,44	10,68	7,05
VvSTS20	1517	29,29	13,2	18,44	10,25	37,62	7,86	35,01	15,85	16,05	16,91	26,57	14,58	15,78	25,7	17,31	13,85	16,25	10,45		15,18	12,92	26,61	24,3	24,03		48,34	56,39	46,47	38,82	41,81	40,01	48,07	54,4	46,54	37,88	48,2		8,36	13,65	13,98	8,52	9,72	18,58	11,38
VvSTS21	1537	20,24	10,24	16,56	9,99	26,55	10,25	25,05	9,79	14,72	11,76	17,74	9,99	11,11	19,45	15,9	9,73	10,65	6,7		12,68	9,92	20,04	15,11	14,72		9,46	15,51	15,44	10,45	8,15	8,61	10,51	8,41	8,94	16,03	17,09		0,52	1,45	0,99	0,59	0,72	1,38	0,46
VvSTS22	1315	21,04	9,02	10,22	3	12,67	2,61	13,44	4,91	7,83	6,3	10,68	5,38	5,45	12,37	8,99	7,76	9,22	6,84		7,76	7,6	11,93	7,83	9,52		4,99	6,45	5,38	6,07	2,92	4,69	4,99	6,45	5,53	9,29	8,83		4,38	6,37	7,07	4,99	5,99	8,76	6,37
VvSTS23	1369	19,99	8,47	12,17	6,2	19,11	4,57	26,72	9,89	11,21	11,44	17,78	10,99	6,64	12,98	11,14	10,55	11,14	5,75		10,48	10,99	16,24	12,98	20,95		20,14	16,75	16,6	15,94	21,99	15,79	21,69	21,32	21,25	18,59	17,93		3,2	4,8	4,87	2,51	2,66	5,39	2,43
VvSTS24	1370	29,78	13,99	13,86	5,16	18,87	6,41	13,41	6,49	15,11	8,77	8,92	5,97	7,15	13,42	12,75	9,51	6,56	6,12		8,04	7,96	13,63	9,36	10,47		20,13	18,06	14,01	11,8	13,71	12,46	14,3	14,97	14,67	14,45	15,26		0,68	1,84	2,21	1,25	0,88	2,51	1,62
VvSTS25	633	18,03	8,6	11,49	7,66	17,23	1,12	20,17	7,5	8,3	7,66	11,97	8,46	6,7	10,05	8,46	7,18	6,86	7,82		6,06	9,09	9,39	9,73	12,45		16,75	18,99	10,21	15,64	17,39	16,12	14,84	21,22	11,65	18,03	8,78		2,51	7,82	6,38	2,71	4,31	9,41	5,27
VvSTS26	1101	51,1	29,65	4,5	2,48	8,9	1,19	13,82	2,84	3,39	4,04	7,98	5,23	5,69	5,87	4,22	3,3	9,72	5,87		17,25	11,47	9,17	10,37	27,34		3,58	3,12	2,66	2,75	1,83	1,93	3,39	2,11	2,02	4,68	3,12		0,9	2,94	2,29	1,19	1,19	4,86	2,2
VvSTS27	1535	50,6	25,02	9,01	5,2	11,78	2,63	13,36	8,36	9,34	6,19	11,78	6,84	6,97	8,88	5,59	5,26	10,73	7,83		13,23	10,07	12,64	9,8	19,21		6,51	7,9	5,79	7,63	9,15	9,08	9,61	7,24	8,16	9,94	5,46		3,39	6,78	5,59	4,28	3,29	6,12	3,36
VvSTS28	1517	47,6	25,33	16,51	8,39	22,77	2,33	25,15	8,46	13,98	11,25	18,64	13,25	11,12	16,31	8,59	9,12	9,19	11,45		17,38	13,98	25,91	16,51	25,43		5,19	5,66	3,6	3,86	4,39	6,19	3,86	4,93	3,53	10,99	10,32		1,89	2	3,86	2,93	2,93	6,13	3,73
VvSTS29	1535	50,67	21,96	19,34	6,58	22,96	2,9	22,97	10,07	13,62	9,28	19,21	11,98	12,96	14,54	11,12	8,42	10,13	12,44		16,71	12,37	24,49	18,42	25,39		14,87	15,07	12,76	15,46	15,07	14,67	11,38	14,28	9,67	20,53	11,05		4,84	8,36	7,96	7,17	5	11,19	5,46
VvSTS30	1533	23,52	7,99	7,44	5,6	13,05	3,29	8,68	6,19	6,65	7,51	11,33	6,92	6,13	7,58	8,63	4,55	7,25	3,29		8,56	5,6	11,94	7,38	10,41		5,47	8,7	6,32	5,53	7,64	6	6,65	6,26	3,29	9,49	6,19		2,55	2,64	3,29	2,31	2,83	4,94	3,75
VvSTS31	1538	22,85	12,34	10,18	6,3	14,32	3,41	12,17	7,16	8,21	8,34	12,87	6,83	7,81	7,95	8,47	5,58	6,96	5,98		9,85	5,98	12,39	8,54	10,44		8,87	12,61	12,08	7,49	10,11	9,26	8,34	9,65	8,54	12,74	13,27		8,62	10,31	13,86	10,44	6,96	14,84	8,47
VvSTS32	1387	25,49	13,34	5,46	3,28	10,92	2,62	9,84	3,2	5,46	4,01	7,14	4,66	4,73	5,24	5,75	3,06	3,71	2,33		8,01	6,99	19,1	6,84	10,19		5,53	8,74	9,61	7,14	6,48	7,06	6,26	7,65	7,57	8,67	6,48		2,28	3,35	2,62	2,48	3,57	5,46	3,79
VvSTS33	1563	27,14	11,56	0,9	1,36	2,26	0,71	5,68	0,58	1,23	1,55	1,74	1,09	1,55	1,42	1,16	0,45	1,42	1,03		5,94	6,14	8,53	4,91	13,89		2,71	3,23	3,61	2,58	3,16	1,49	3,03	3,17	2,45	2,32	1,94		0,1	0,13	0,26	0,19	0	0,32	0,06
VvSTS34	453	18,95	9	3,79	2,68	7,13	1,56	2,68	1,34	4,01	3,12	4,01	2,23	1,34	0,67	2,68	2,23	1,56	1,78		4,24	3,12	8,99	0,89	7,58		0,22	0,22	0	0,45	0	0	0,22	0	0	3,34	1,34		0	0	0	0	0	0	0
VvSTS35	1537	23,26	9,39	11,23	4,93	15,05	4,27	10,88	7,88	7,75	7,49	10,18	5,45	6,97	8,67	7,23	5,52	5,19	6,24		9,4	7,42	10,48	7,62	12,16		3,42	3,29	5,45	2,1	3,15	3,61	2,43	2,83	2,43	6,7	6,04		1,57	2,23	1,12	1,12	0,59	2,1	0,92
VvSTS36	1491	23,37	11,71	11,52	6,16	18,22	4,88	14,09	9,89	11,11	9,55	13,95	9,01	6,44	11,38	11,45	6,23	6,16	6,71		11,65	9,28	12,09	11,52	12,74		9,48	13,75	12,06	9,28	10,29	10,7	8,54	7,52	5,62	10,3	11,38		2,82	5,08	4,95	3,79	3,93	7,65	4,67
VvSTS37	1538	18,78	10,68	10,38	5,71	15,7	2,63	11,25	7,22	8,6	7,49	11,89	8,54	7,29	9,33	7,29	6,3	6,96	5,71		9,19	8,54	11,4	9,78	12,54		18,39	15,96	24,3	10,97	14,25	13,2	11,49	9,13	12,15	12,54	10,9		4	7,75	6,17	3,61	2,76	7,88	3,61
VvSTS38	1555	23,32	9,92	8,64	5,78	14,35	3,38	11,36	7,79	8,96	6,3	15,2	8,12	7,27	6,88	8,7	3,7	8,9	5,78		8,64	7,08	14,2	8,77	12,34		20,46	20,39	26,05	10,59	16,63	17,21	16,11	15,13	12,73	15,2	13,12		4,55	8,38	7,53	4,74	4,42	8,31	6,17
VvSTS39	1557	23,35	10,2	6,94	5,25	13,1	2,21	14,08	5,06	8,3	6,23	10,51	6,42	4,54	8,24	5,12	4,54	7,72	3,89		8,24	7,2	11,48	8,56	11,48		7,92	7,85	10,06	6,75	7,66	6,62	6,43	4,09	5,65	7,59	7,59		3,26	5,97	3,57	4,02	2,46	5,71	2,85
VvSTS40	1555	26,05	10,31	3,7	4,29	8,7	1,62	6,74	4,94	4,61	3,18	9,09	6,43	3,7	3,12	3,7	1,62	4,42	2,01		11,56	5,46	5,27	3,96	8,38		0,91	0,71	1,43	0,58	0,52	0,84	0,52	1,36	0,32	5,26	5,91		0,03	0	0	0	0	0	0
VvSTS41	1539	23,04	9,03	5,78	8,07	10,63	2,03	12,54	9,32	7,88	6,5	14,44	6,43	7,28	5,78	7,88	3,35	8,66	2,76		10,7	7,28	8,04	8,53	8,47		5,84	6,76	8,79	5,12	5,05	6,3	5,12	6,56	4,86	13,06	8,73		2,07	3,41	2,03	2,76	1,77	4,33	2,43
VvSTS42	1556	20,64	11,45	4,54	4,87	10,52	2,08	12,31	5,65	6,17	7,4	8,89	6,49	5,52	4,61	4,93	2,14	7,72	3,12		7,79	5,58	9,13	7,72	9,15		2,66	3,57	1,36	2,99	3,57	2,47	1,82	3,89	2,47	6,3	4,8		1,14	2,01	2,21	1,43	1,36	1,82	1,17
VvSTS43	1558	22,24	8,03	1,36	1,62	1,69	0,26	1,41	1,1	0,97	1,17	1,62	4,02	0,84	1,04	1,1	0,65	1,3	0,39		4,28	4,54	1,91	1,36	7,46		15,5	14,13	17,19	11,41	13,81	10,38	11,35	9,01	10,5	8,3	7,39		1,83	3,76	1,69	1,82	2,14	4,02	1,62
VvSTS44	1555	21,63	10,9	3,64	5,33	10,13	1,36	8,31	5,78	5,46	5	10,39	7,27	4,42	4,74	4,16	2,34	4,68	2,99		9,29	5,91	5,05	5,26	10,59		1,62	1,62	2,34	0,52	1,23	0,52	0,78	2,66	1,49	3,25	4,09		0	0,06	0,06	0,13	0,06	0,06	0
VvSTS45	1539	22,9	9,5	3,02	2,89	5,18	0,72	12,93	4,07	2,76	3,87	7,81	10,76	3,48	2,36	4,13	1,77	9,71	1,12		7,61	5,32	3,73	9,06	11,68		2,17	3,08	3,54	2,69	3,08	3,15	2,43	4,46	2,49	4,79	3,54		0,94	1,51	1,64	1,18	0,98	3,02	1,18
VvSTS46	1556	24,28	8,39	7,79	2,73	9,09	2,53	7,69	2,34	6,75	4,15	5,19	3,38	2,14	6,88	6,1	4,74	4,22	4,35		6,17	6,82	13,38	4,02	9,8		3,89	3,63	1,17	2,79	3,38	1,88	3,18	3,05	2,08	3,57	3,57		1,21	2,27	1,75	1,36	1,49	2,53	0,91
VvSTS47	1540	23,22	10,51	14,95	4,33	16,85	4,33	6,51	5,9	10,36	10,49	9,12	5,12	5,84	14,43	13,05	10,3	5,12	9,58		7,28	6,82	24,38	5,12	11,94		5,84	7,28	9,12	5,51	6,76	6,89	6,1	6,49	6,89	7,87	7,35		2,63	2,56	3,08	1,84	1,25	3,8	1,97
VvSTS48	1538	21,01	10,65	15,56	5,45	23,05	4,4	9,9	9,13	13,27	13,33	13,46	8,14	9,46	13,79	13,86	7,29	6,3	7,95		11,16	7,36	18,06	8,73	10,7		8,54	9,92	11,69	9,72	9,98	7,49	7,95	11,03	6,83	12,54	10,38		2,14	4,2	3,22	1,9	1,97	4,66	2,3

Row represents STS Genes (1 – 48)
Column represents Genotypes (each block – different category of species)

Red: Remarkably high coverage
Yellow: much closer to the values of high coverage
Blue: Very low or completely missing reads

CLC Genomics Workbench

A comprehensive and user-friendly analysis package for analyzing, comparing, and visualizing next generation sequencing data.



Map Reads to Reference

1. Select sequencing reads
2. Set references
3. Set mapping parameters

Set parameters

Selected reads

	Input	Length	Type	Settings
	Regale-Truseq_C0RMCA...	Long	Paired	User setting

Long reads mapping parameters

Mismatch cost 2

Insertion cost 3

Deletion cost 3

Length fraction 1,0

Similarity 0,99

☐ Global alignment

☐ Color space alignment

Colorspace error cost 3

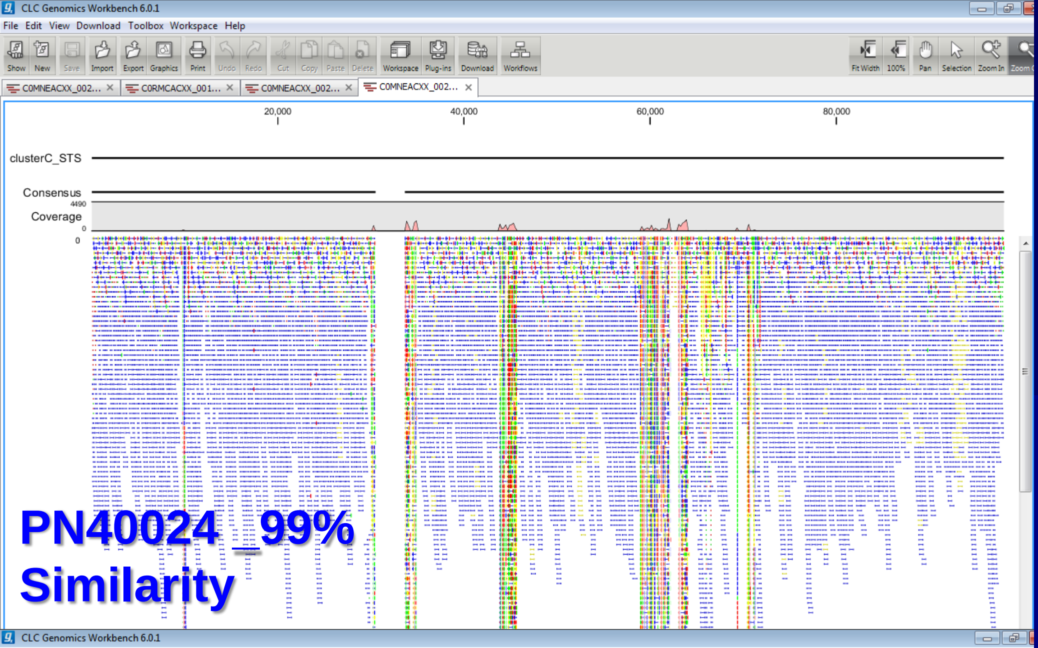
Paired parameters

☐ Override distance setting

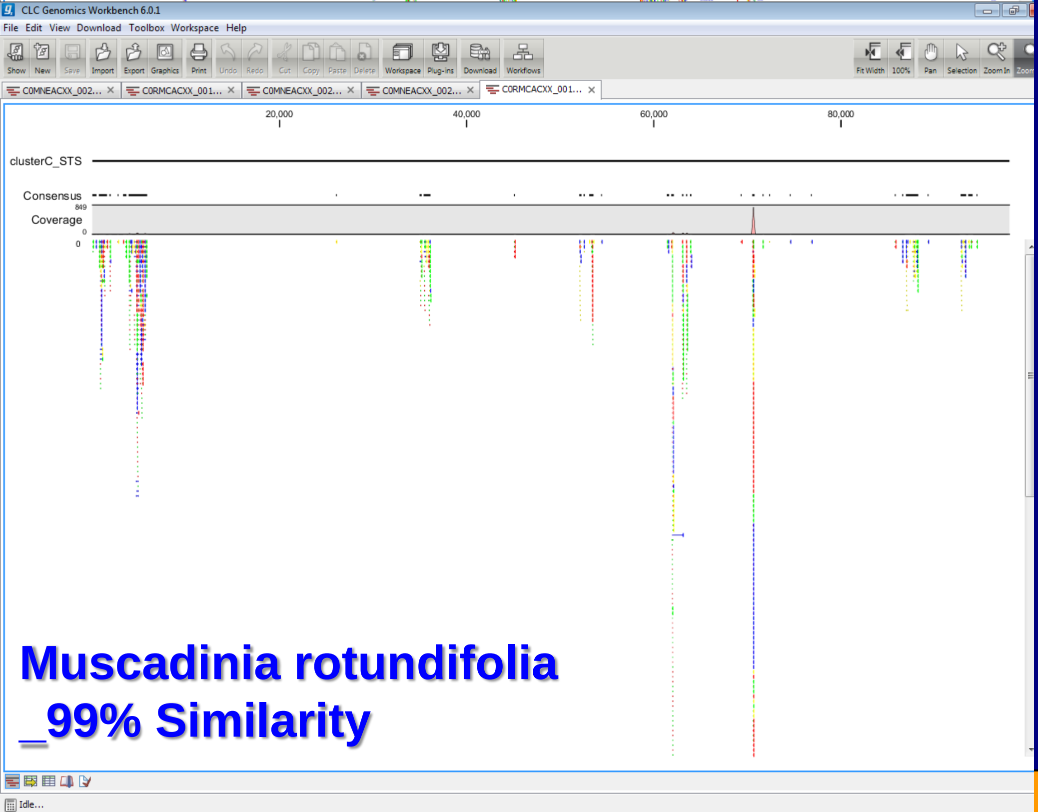
Minimum distance 180

Maximum distance 550

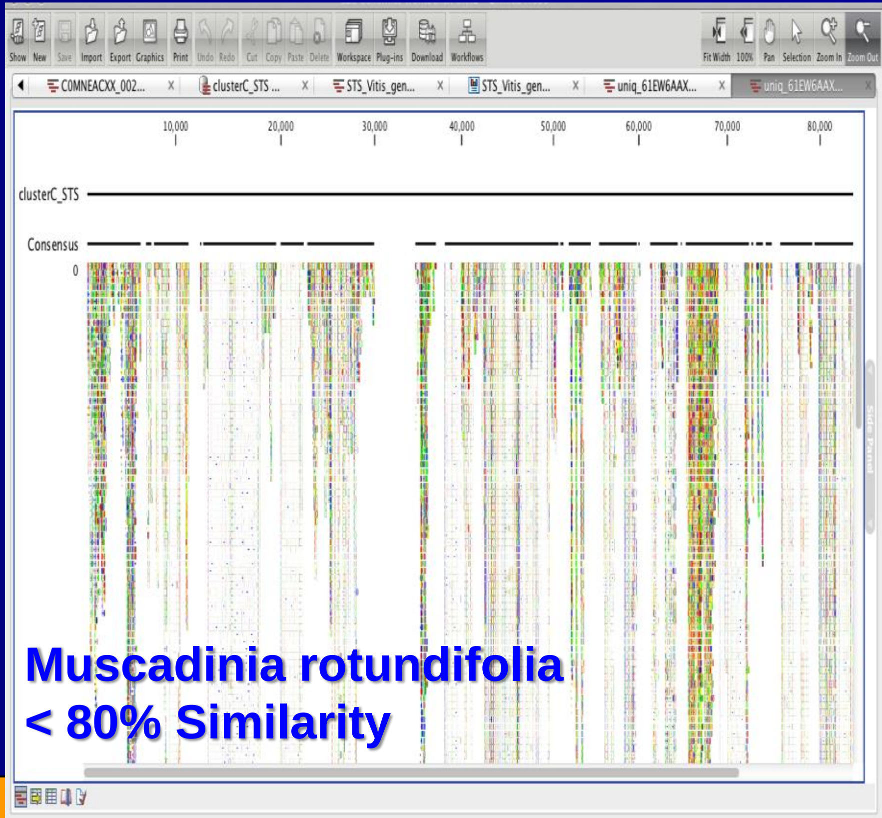
STS – Cluster C



PN40024 _99% Similarity



Muscadinia rotundifolia _99% Similarity



Muscadinia rotundifolia < 80% Similarity

Annotation of RGA genes in PN40024

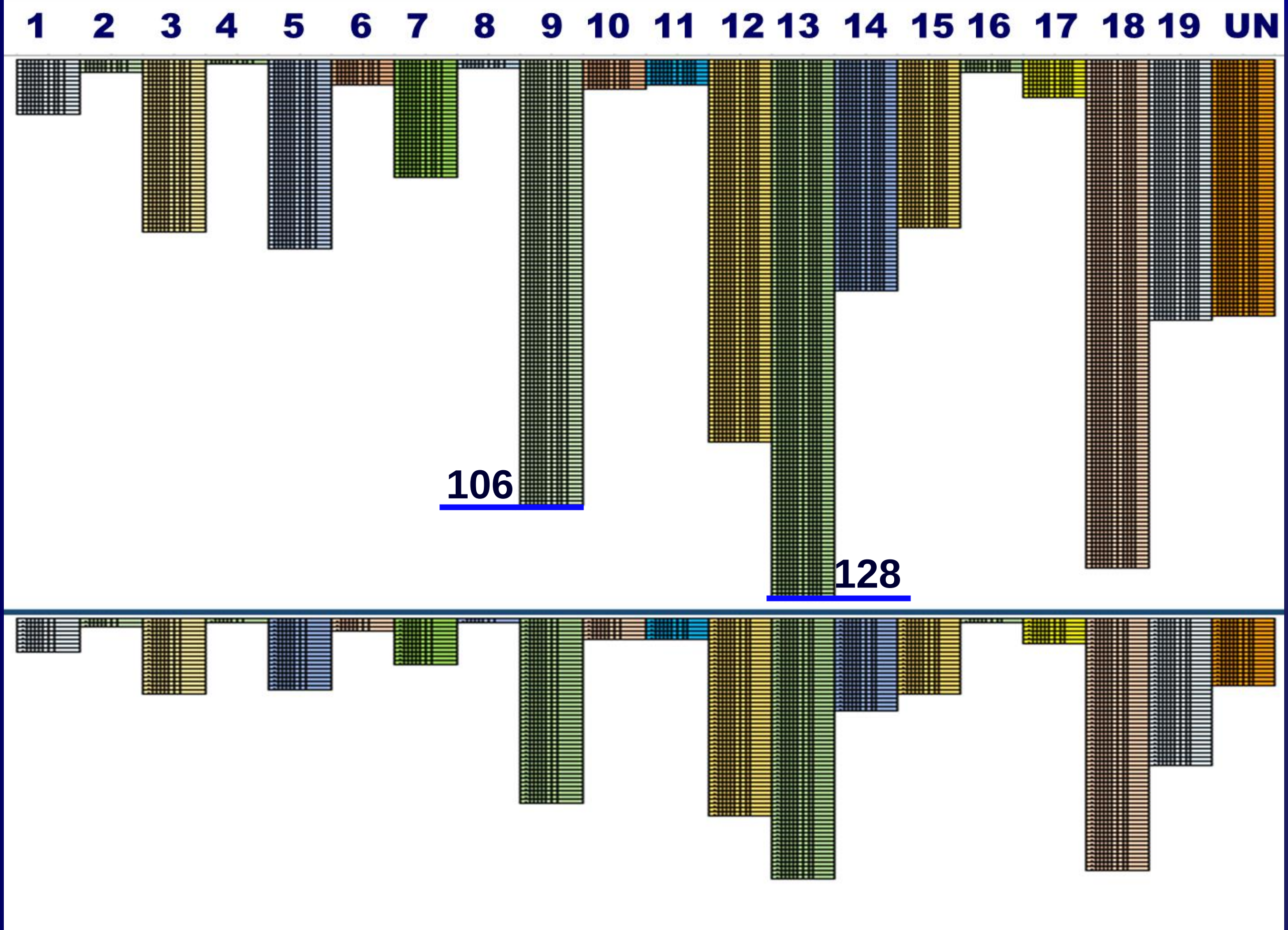
URGV - Plant Genomics Research
Unité de Recherche en Génomique Végétale

- The 12x Vitis vinifera (PN40024) genome has been explored exhaustively using a similarity search approach based on a reference bank made with previously characterized NBS-LRR proteins.

- A total of 828 NBS-LRR genes were identified and characterized in PN40024.

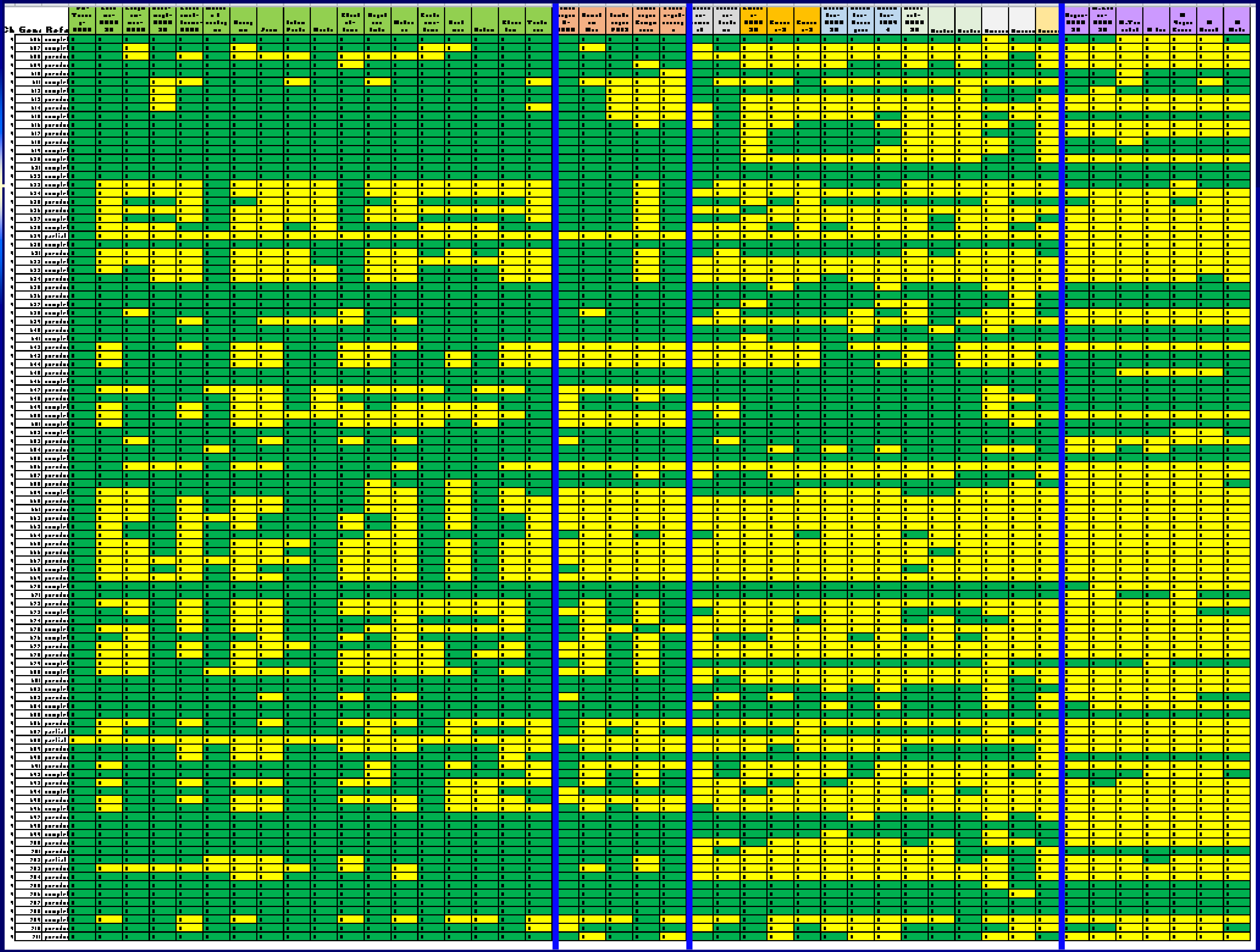
- These 828 NBS genes are chosen to study their variations, presence-absence among the genotypes.

Global view of RGAs scattered across the 19 chromosomes of PN40024

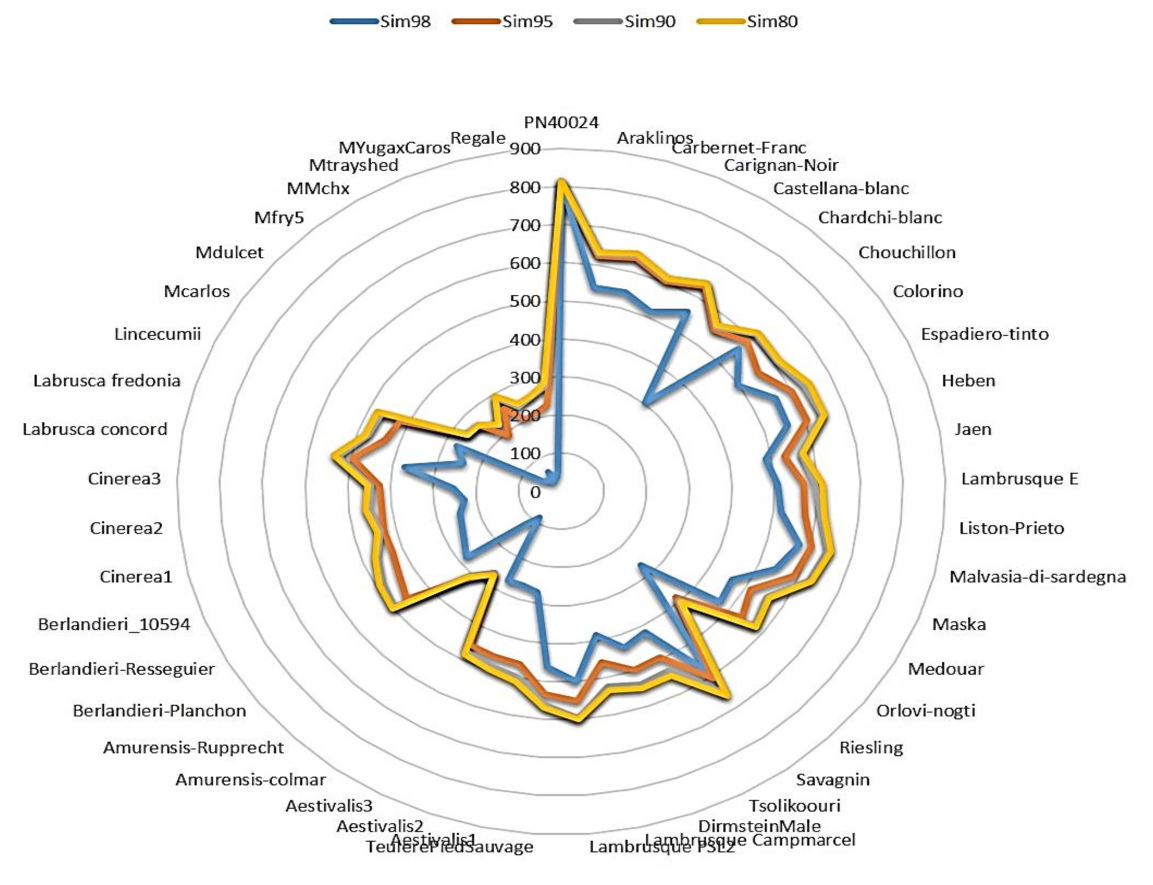


Presence and absence of genes (based on reference length coverage)

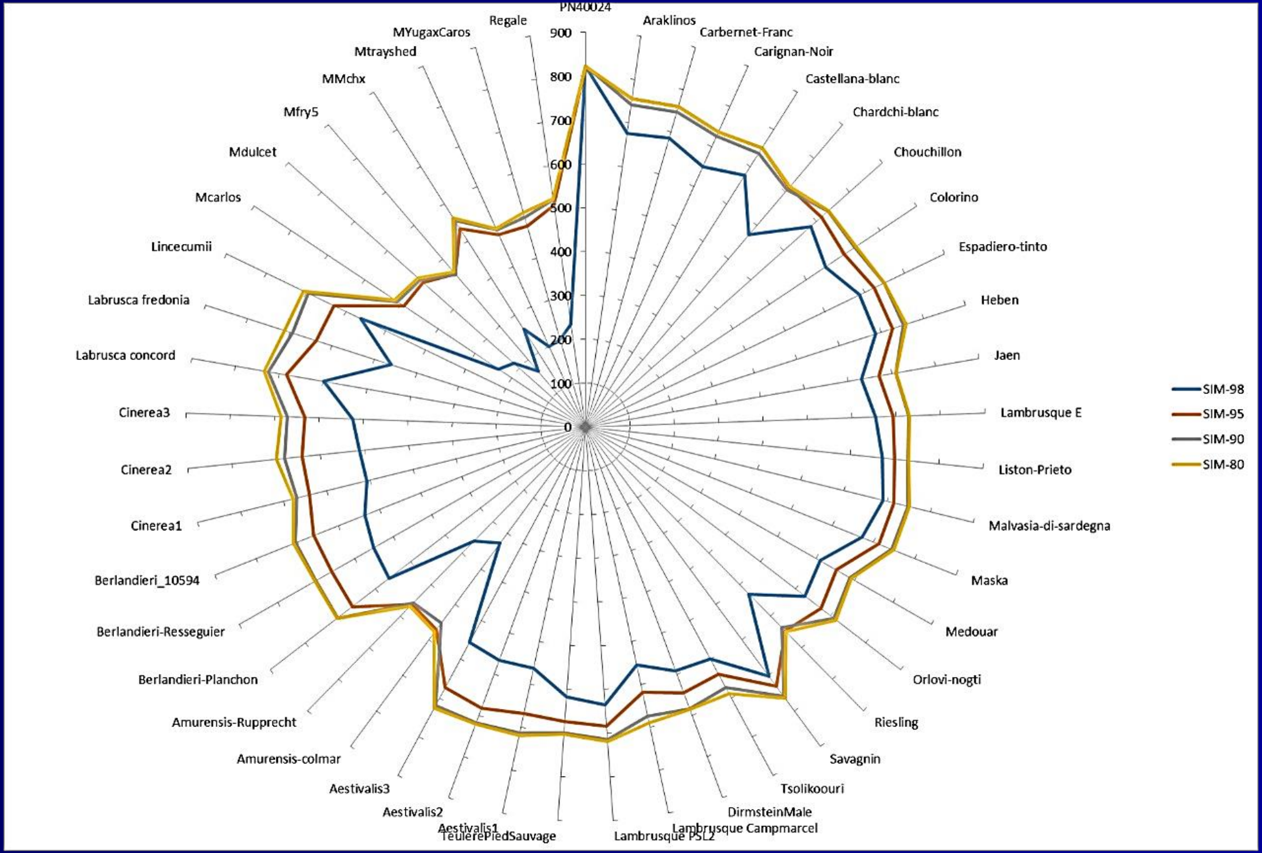
Chromosome 9



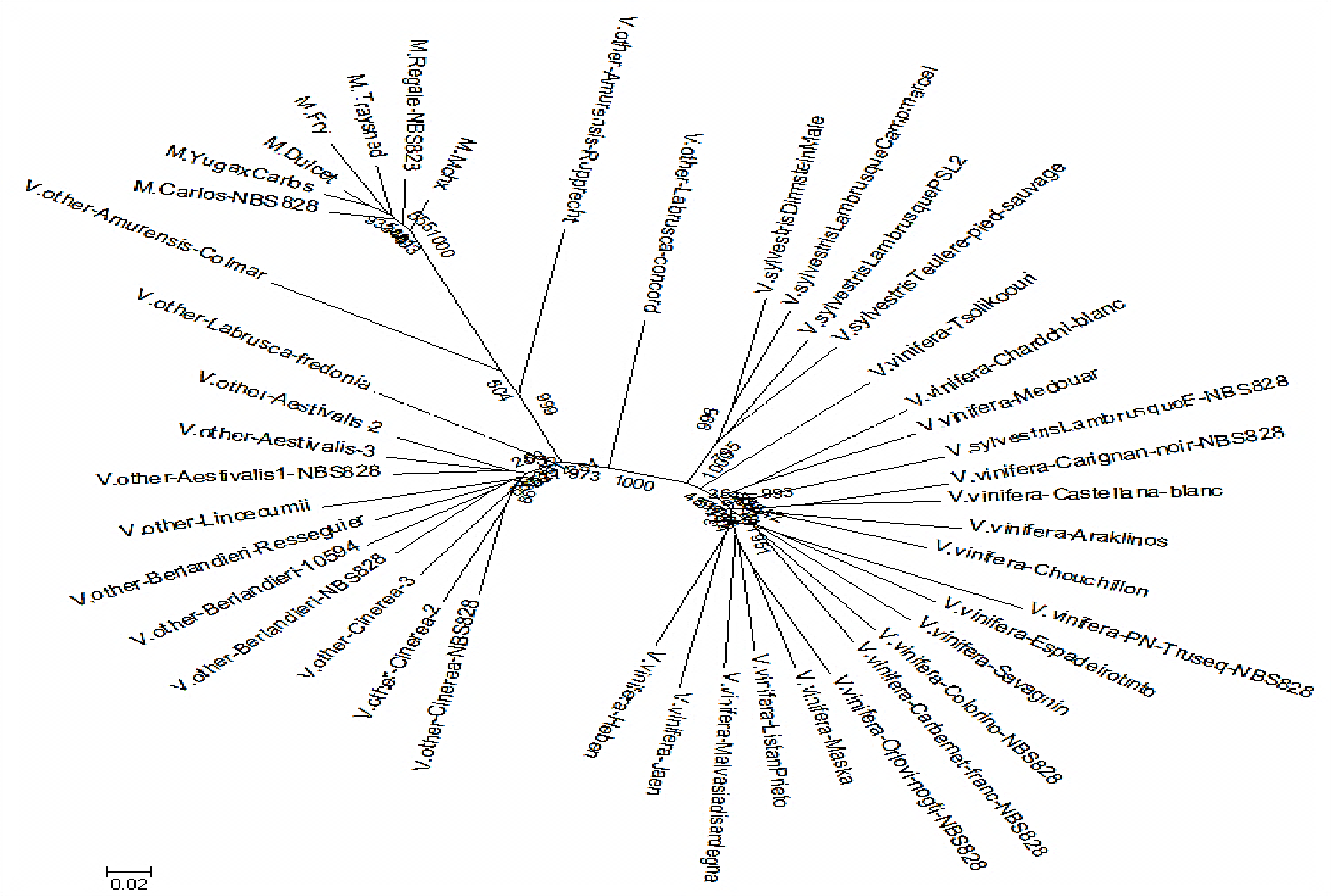
Reference genes covered by length over 90%



Reference genes covered by length over 65%



NJ tree depicting the phylogenetic structure of the Vitis and Muscadine genotypes



Limitations and Challenges

- Only Reference genes can be compared
- Species specific haplotypes / gene sequences are ignored
- Tedious validation
- Heterozygosity - creating false segmental duplications in assemblies.

Perspectives

- **Local Assembly based strategy**
- **Longer-reads**
- **Motif-based mapping strategy**

In-search of ideas and solutions

GenScale

Scalable Optimized and Parallel Algorithms for Genomics

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche RENNES - BRETAGNE ATLANTIQUE

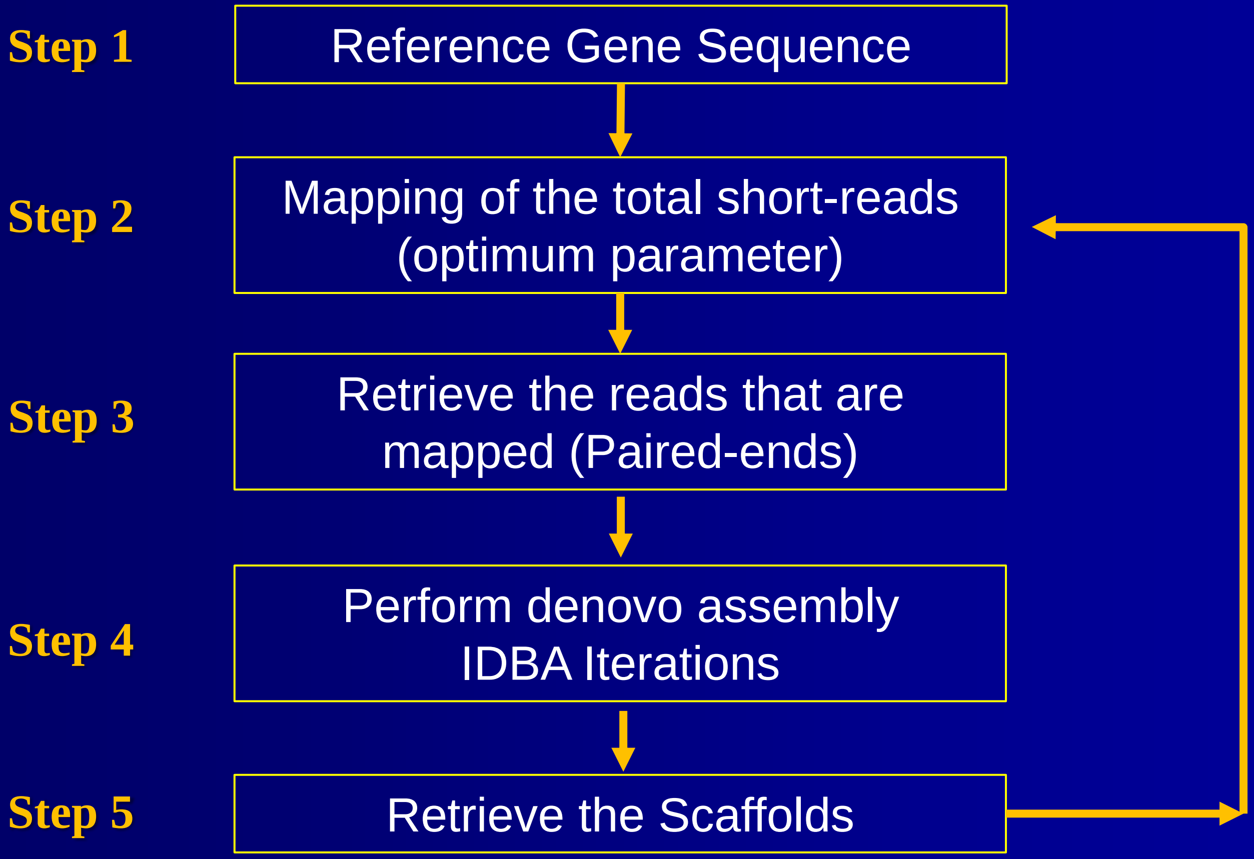
- Local softwares and algorithms didnt meet our requirements
  - Paired-end reads
  - Large data sets
  - Longer-reads??

hku-idba

A Practical Iterative de Bruijn Graph De Novo Assembler

Customized CLC-IDBA Iteration approach

- Based on Christelle Aluome’s experiences (at EPGV) in Genespace (Pea)



Step 1 Reference Gene Sequence

Step 2 Mapping of the total short-reads (optimum parameter)

Step 3 Retrieve the reads that are mapped (Paired-ends)

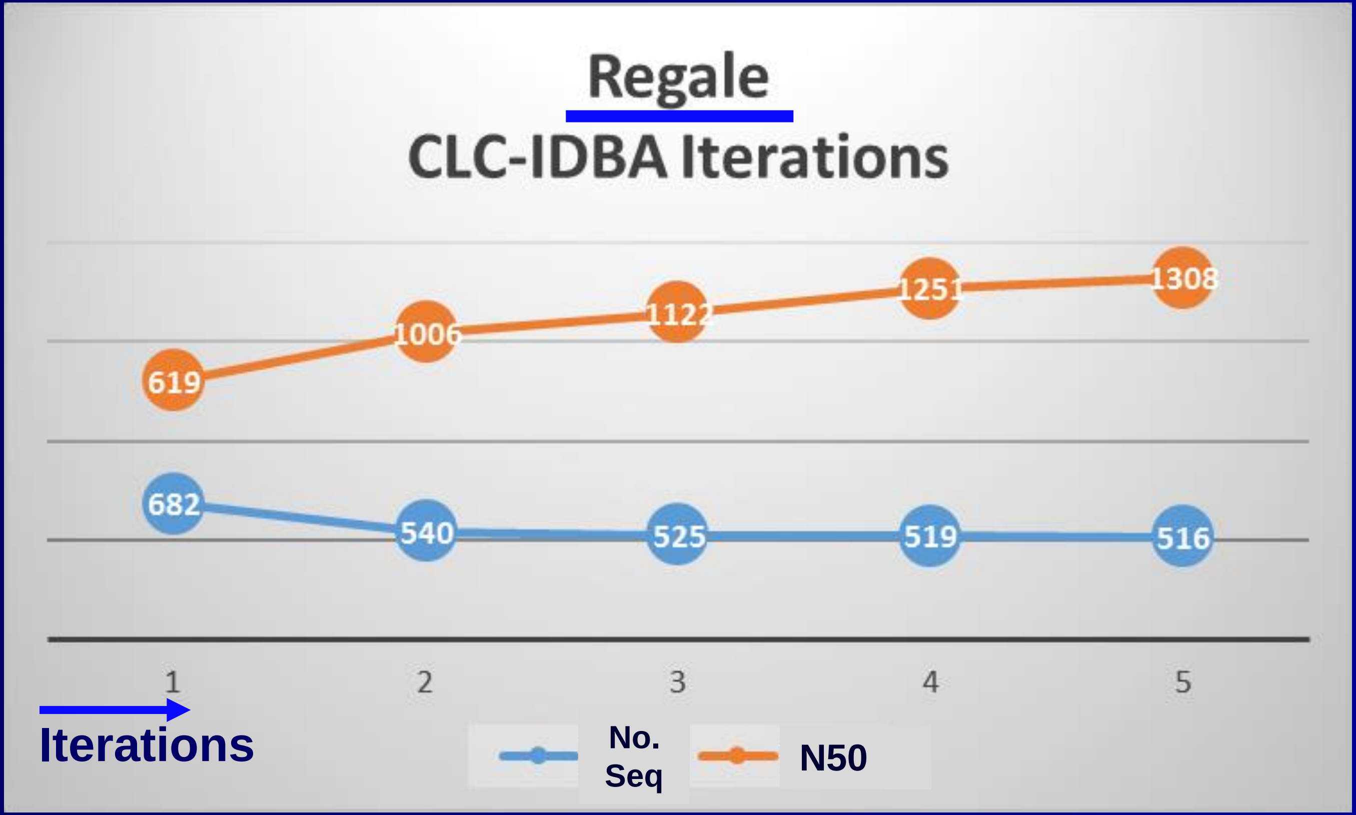
Step 4 Perform denovo assembly IDBA Iterations

Step 5 Retrieve the Scaffolds

In Progress…

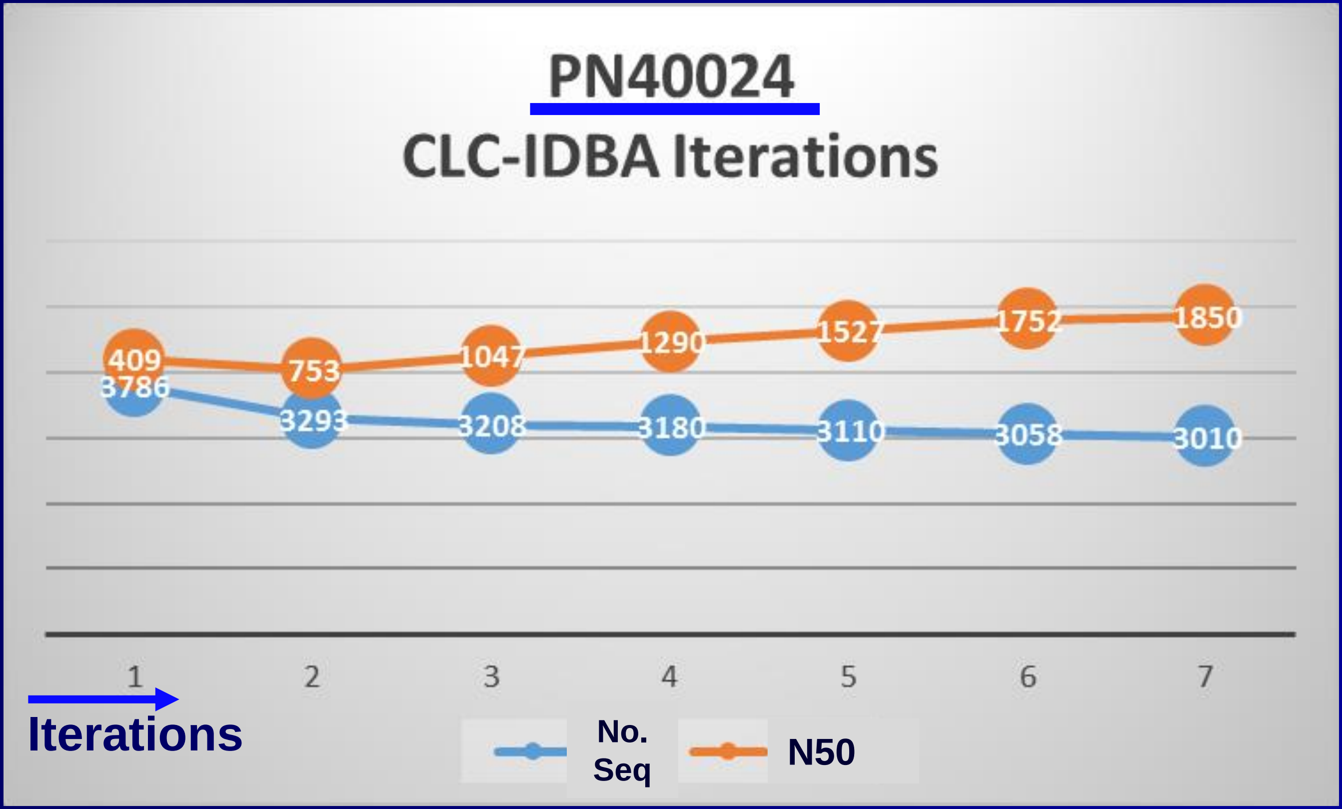
(hopeful signs..)

828 RGAs



In Progress…

(hopeful signs..)



Conclusion

- **Rigorous increase in sequencing scales - Paucity of tools to analyse resequenced data in regards to gene families**
- **Need more improvement in species-specific haplotypes and gene discoveries**
- **Efficient Local assembly tools – need of the hour!**

Acknowledgements

INRA - Evry
Dominique Brunel
Marie-Christine Le-Paslier

EPGV
Etude du Polymorphisme des Génomes Végétaux

INRA - Evry
Sebastien Aubourg
Nadia Bentahar
Cecile Guichard
Marie-Laure Martin Magniette
Anne-Françoise Adam-Blondon

Unité de Recherche en Génomique Végétale
URGV

INRA - Colmar
Camille Rustenholz
Sophie Blanc
Pere Mestre
Emilce Prado
Philippe Hugueney
Didier Merdinoglu

Santé de la Vigne et Qualité du Vin (SVQV)

Thank You !