# ConquestExplorer



## A Genomic and Phenotypic Data Repository and Processing Tool towards Omics Data

Rico Basekow<sup>1</sup>, Jost Neigenfind<sup>1</sup>, Axel Nagel<sup>1</sup>, Christiane Gebhardt<sup>2</sup>, Birgit Kersten<sup>1</sup> **Contact:** basekow@mpimp-golm.mpg.de

<sup>1</sup>GabiPD Team, Bioinformatics, Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14424 Potsdam-Golm, Germany <sup>2</sup>Max Planck Institute for Plant Breeding Research, Carl von Linné Weg 10, 50829 Cologne, Germany

#### Introduction

The GABI-Papatomics project is aiming at the validation of the diagnostic power of specific marker allele combinations for resistance to late blight and tuber quality traits in marker-assisted selection (MAS) experiments in populations of tetraploid potato genotypes. The used markers are derived from the former Gabi projects Gabi-Gabi-Chips. Conquest and metabolome transcriptome, proteome and MAS between sub-populations differences contrasting genotypically and phenotypically for resistance to late blight will be analysed.

#### Data Repository

The ConquestExplorer manages data in a relational database (Fig.1) which is accessed by an objectrelational data layer. Thereby the underlying data model is flexible, easy expandable and adaptable to new data types and requirements. The tool allows to import data from files as well as to insert and to manipulate data on forms. To simplify data input a Furthermore new input form similar to a spreadsheet was developed. Data can be explored by navigation and retrieved by a full-text search. Furthermore, a selection tool for marker profiles (Fig.2) was developed to support MAS.

#### Data Analysis

The integration of data management and data analysis makes data processing easier, more transparent and reproducible compared to a "manual" data analysis with a lot of cumbersome intermediate steps like data reformatting. Currently, standard statistical tests are integrated like Anova and Kruskal-Wallis (Fig.4). That enables correlation analyses between genotypic and phenotypic traits. The ConquestExplorer does also have an interface to the SATIotyper, a tool to predict haplotypes from unphased SNP data from heterzygous polyploids [2,3].

Efficient managing and processing of data are important tasks for any project which handles huge amount of data. ConquestExplorer is a data repository for genomic and phenotypic data. The tool provides an easy and well arranged access for biologists and breeders to their data.

### ConquestExplorer

ConquestExplorer stores the data in a backend database, provides a graphical user interface (GUI) to maintain the data and has an interface to the statistic software R to analyze data. Currently, we manage data of the Conquest2 project and initial data of the Papatomics project with this tool. **Figure 2**: ConquestExplorer – Data management Conquest2 data consist of more than 4,000 PCR Data Visualization products with sequence information and tracefiles, about 45,000 SNPs, 52,000 PCR markers and 17,000 phenotypic data points. The current Papatomics data contain more than 13,000 different types of project data, the tool supports marker data points of about 1,600 genotypes. miscellaneous visualizations. There are text, map and image views. Those can show "raw" data like sequences or "correlated" data like SNPs that will be speciesId genus (FK) (Fig.3). strainId species ( \_ **D** X ConquestExplorer-Conquest Genotype genotypelo name <u>F</u>ile <u>M</u>odel <u>H</u>elp strain (FK ploidie name genotype submissionId (FK) ioobjectId mapName ubmissionId (Fl tissueName (FK species 🗖 🗖 SequencerSequence - CONQUES... 🕱 📄 SequencerSequence - CONQUES.. 🐕 Model 🍃 Search genotype (FK) hromosomenN napUnit description name uc Feature (123061) vector developme rsite1 👂 🗟 Genotype (229) Greencards Sequence: 239E4left Con-B10:10-L032088-099 rsite2 libname 🕨 🗟 Locus (81) view host antibiotic trace length: 1099 IementName mapId (FK) 🕨 🗟 Map (36) accattatte ageceattte atatgaaatt gttatget agaatettt tTatatttat bioobjectId (FK) ubmissionId (FK) gaggaggg tttcctgGga agggtac cc Gtcgaacaag aatgtccCGt 131 hromosomalPositio Clone tettGttte aagtagttga ggaattTtat eteattgttt gaaeTtttTg 191 bioobjectld (F 251 e e ac gagtgtttgt tgateetgae aetgeaagaa atetgttttt t<mark>T</mark>eeaggaat omalPosition 311 atttgateaa gagaetgetg eagttgtagt ggetagatta gettettaea libraryld (FK) wellPlate 331 Genetic wellRow wellColumn F1840 sequenceld equencedObject (Fl left flanking [allel] right flanking map fullLength ubmissionId (FK orimer (FK) 39E4left Con-B10 INDEL84 agaatetttt[T/T/T/T]atatttattt PcrProduct bioobjectld (FK) percentages eversePrimer (F sequenceDirection forwardPrimer (FK sequenceLob submissionId (FK primer (FK) name type eneFunction 239E4left Con-B10 SNP133 agggtacete[G/G/G/G]tegaacaaga sequenceDescript name\_uc ubmissionlo geneld geneFunctionId abbreviation sequence\_snpl name 🚂 🚮 🛃 🛃 🗖 🖬 iole X

🍃 Mo 🛛 🍃 Se		∃ *Datashee	t-MapElement	(0) 🖂					- 8
	\$	id	colou	r name	positionStart	positionStop	mappedMaterial	map	<b>A</b>
Feature (126,169)		0 411	magent		4.003		BA202k7 4884	#1	
Genotype (229)		1 25		BA31n11	10.004		PA21011 4624	#1	_
		2 1116		BA219k9	13.004	Co	ontent <sup>22</sup>	#1	
🕨 🗀 Locus (81)		3 1129		BA213c14	15.001		/iew 718	#1	- 11
🕨 🧀 Map (36)		4 144		BA228g19	37.0		10	#1 #1	- 11
▽ 🗀 MapElement (1,334)		5 29 6 1117		BA1m23 BA27c1	10.008 13.005	7/	BA27c1 5303	#1	- 11
		7 510		BA2701 BA30p15	55.004		BA30p15 4946	#1	- 11
Navigation		7 510		BA30015	Feature Selecti		BA30013 4940	#1	
At1A - At1B (11)	genotypeS	Set: SC						selected/all:	576/1
	genotypeS Selection	Set: SC						selected/all:	576/1
At1B - At1G (11)	Selection		featureNa		alleleValue:		eDosage:	count:	
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> </ul>	Selection		featureNa			allel			576/1
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> <li>At2 - At2 (11)</li> <li>At2B - At2E (11)</li> </ul>	Selection			) -	0		leDosage:	count:	dele
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> </ul>	Selection locusName GP171		- a (600 bp	) -		•	leDosage: Marke	count: 76 r profile	dele
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> <li>At2 - At2 (11)</li> <li>At2B - At2E (11)</li> </ul>	Selection locusName GP171 StpL StpL		<ul> <li>a (600 bp</li> <li>3b (1236</li> <li>3e (360 bp)</li> </ul>	) -		•	leDosage: Marke	count:	dele
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> <li>At2 - At2 (11)</li> <li>At2B - At2E (11)</li> <li>At2E - At2F (11)</li> </ul>	Selection locusName GP171 StpL		<ul> <li>▼ a (600 bp</li> <li>3b (1236</li> </ul>	)	0 0 1 1	•	leDosage: Marke	count: 76 r profile	dek -
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> <li>At2 - At2E (11)</li> <li>At2B - At2E (11)</li> <li>At2E - At2F (11)</li> <li>A - B (11)</li> </ul>	Selection locusName GP171 StpL StpL Stp23		<ul> <li>a (600 bp</li> <li>3b (1236</li> <li>3e (360 b)</li> <li>8b (348 b)</li> </ul>	)	0       0       1       1       0		leDosage: Marke	r profile	
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> <li>At2 - At2 (11)</li> <li>At2B - At2E (11)</li> <li>At2E - At2E (11)</li> <li>At2E - At2F (11)</li> <li>A - B (11)</li> <li>A - B (11)</li> <li>BA11 - BA18 (11)</li> </ul>	Selection locusName GP171 StpL StpL Stp23 Rca InvGE		<ul> <li>a (600 bp</li> <li>3b (1236</li> <li>3e (360 b)</li> <li>8b (348 b)</li> <li>1a (~300</li> <li>6f (368 b)</li> </ul>	)	0         0         1         1         0         1         1         1         1         1         1         1		leDosage: Marke	r profile	dele 
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> <li>At2B - At2E (11)</li> <li>At2B - At2E (11)</li> <li>At2E - At2F (11)</li> <li>A - B (11)</li> <li>A - B (11)</li> <li>BA11 - BA18 (11)</li> <li>BA1 - BA3 (9)</li> </ul>	Selection locusName GP171 StpL StpL Stp23 Rca InvGE AGPb	e:	<ul> <li>a (600 bp</li> <li>3b (1236</li> <li>3e (360 b)</li> <li>8b (348 b)</li> <li>1a (~300</li> </ul>	)	0         0         1         1         0         1         1         1         1         1         1         1		leDosage: Marke	r profile	
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> <li>At2B - At2E (11)</li> <li>At2E - At2E (11)</li> <li>At2E - At2F (11)</li> <li>A - B (11)</li> <li>A - B (11)</li> <li>BA11 - BA18 (11)</li> <li>BA1 - BA3 (9)</li> <li>B - C (164)</li> </ul>	Selection locusName GP171 StpL StpL Stp23 Rca InvGE AGPb	e:	<ul> <li>a (600 bp</li> <li>3b (1236</li> <li>3e (360 b)</li> <li>8b (348 b)</li> <li>1a (~300</li> <li>6f (368 b)</li> </ul>	)	0         0         1         1         0         1         1         1         1         1         1         1		leDosage: Marke	r profile	
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> <li>At2B - At2E (11)</li> <li>At2E - At2E (11)</li> <li>At2E - At2F (11)</li> <li>A - B (11)</li> <li>A - B (11)</li> <li>A - B (11)</li> <li>A - BA11 - BA18 (11)</li> <li>BA11 - BA3 (9)</li> <li>B - C (164)</li> <li>C - G (127)</li> <li>GP1 - GP2 (130)</li> </ul>	Selection locusName GP171 StpL StpL Stp23 Rca InvGE AGPb	e:	<ul> <li>a (600 bp</li> <li>3b (1236</li> <li>3e (360 b)</li> <li>8b (348 b)</li> <li>1a (~300</li> <li>6f (368 b)</li> <li>10a (501</li> </ul>	)	0         0         1         0         1         0         1         0         1         0         1         0         1         0         1         0		leDosage: Marke	r profile	
<ul> <li>At1B - At1G (11)</li> <li>At1 - At2 (11)</li> <li>At2B - At2E (11)</li> <li>At2E - At2E (11)</li> <li>At2E - At2F (11)</li> <li>A - B (11)</li> <li>A - B (11)</li> <li>BA11 - BA18 (11)</li> <li>BA1 - BA3 (9)</li> <li>B - C (164)</li> <li>C - G (127)</li> </ul>	Selection locusName GP171 StpL StpL Stp23 Rca InvGE AGPb	e:	<ul> <li>a (600 bp</li> <li>3b (1236</li> <li>3e (360 b)</li> <li>8b (348 b)</li> <li>1a (~300</li> <li>6f (368 b)</li> <li>10a (501</li> </ul>	)	0         0         1         0         1         0         1         0         1         0         1         0         1         0         1         0		leDosage: Marke	r profile	

To get a fast and transparent overview about the

							Haplotypes				
									related with phenotyp		
			Locus	PhenoFe	ature _	Pheno' 🕶	Genotype	Haplocypy		r apiocypes	riapioc
			StSGT1-1	Phytoph	thora infestans_2004	160.0	Con-B67	CAATCITCA	CAATCTTCA	CAATGTTCG	CAATG
C		StSGT1-1	Phytoph	thora infestans_2004	161.5	Con-B11	CAATCTTCA	CAATCTTCA	CAATGTTCG	TAATG	
		StSGT1-1		thora infestans_2004	162.5	Con-B28	CAATCTTCA	TAATCTTCG	TAATGTTCA	TAATG	
<u>F</u> ile <u>M</u> odel <u>H</u> elp		StSGT1-1		thora infestans_2004	167.0	Con-B76	CAATCTTCA	CAATGTTCG	CGTCCCCTA	TAATC	
B Cale Cale Control Co			StSGT1-1		thora infestans_2004	167.5	Con-S21	CCAATCTTC	CCAATCTTC	CCAATGTTC	
			StSGT1-1 StSGT1-1		thora infestans_2004 thora infestans_2004	169.0 169.5	Con-B77 Con-B73	CAATCTTCA CGATGTCTG	CAATGTTCG	CGTCCCCTA TAATGTTCG	TAATG TAATG
			StSGT1-1		thora infestans_2004	176.5	Con-S88	CCAATCTTC	CCAATCTTC		CTAAT
▽ 🕞 Feature (123061)			StSGT1-1		thora infestans_2004	178.0	Con-B30	CATCGCCTG	CATCGCCTG	TAATGTTCA	TGATO
<ul> <li>✓ GenotypeFeature (104967)</li> </ul>			StSGT1-1			179.0	Con-S33	CCAATCTTC	CCATTCTTC	CTAATGTTC	TCAAC
			StSGT1-1			179.0	Con-B34	CAATCTTCA	CGTCCCCTA	TAATGTTCG	TAATG
	GenotypeFeatureMa	rker (5225	StSGT1-1	Phytoph	thora infestans_2004	179.5	Con-S30	CCAATCTTC	CCAATCTTC	CCAATGTTC	CTAAT
Þ 🕞	SequenceFeature (5	2711)	StSGT1-1	Phytoph	thora infestans_2004	180.0	Con-S53	CCAATCTIC	CCAATCIIC	CTAATGTTC	СТААТ
			StSGT1-1	Phytoph	thora infestans_2004	183.0	Con-S65	Stati	stics	CCAATGTTC	СТААТ
	✓ ⊂ LocusFeature (608) ▷ ⊂ LocusHaplotypeFeature					Statisti	с	dia	loa		X)AAT
			genotype name (#120):		Con-S				.eg		AAC
▼ PhanatypaEasture (1749			phenotype feature (#1):		Maturity_2006					-	
		phenotyp	phenotype reactive (#1).		Maturity_2006						
		genotype	feature ty	ype:	🖲 SNP 🔾 Hap	lotypeLo	ci 🔘 PCF	Marker∨alu	ie 🔘 PCRMa	arkerDosage	Ca
▷ □ 2005		genotype	/pe mask feature: 🛛 📵 no mask 🔘 1 not mask								
Þ 🗀 2006		genotype	genotype mask feature: 💿 no mask 🔾 1 not mask								
Genotype (229)		genotype	genotype feature (#20): StSGT1						-		
		statistic (	statistic (#3): Anova						-		
👂 🗟 Map	(36)	significan	ce level:		0.001						-
🛛 🕨 🕞 Mapi	Element (133 Res	sults									
N Co sand	()	ew						Ca	lculate	Cancel	
											as
	IsPart PhenoTrait	Signifi			AlleleFrequence		enoMin /	AllPhenoMax	AllPhenoAver	nam	e as
StAOS2 SNP		-	1873360423	311E-4	(120/120 gF: 0 ∨F: 0	) 30.0		217.0	142.4		
D 4 - 0		-	6867670539		(94/120 gF: 26 vF: 0			212.5	146.2287		•
	7 Starch_2004	8.6713	3598794169	02E-4	(96/120 gF: 24 vF: 0	) 10.1		25.4	17.1875		
		0004 0.005	7700297891	73E-4	(120/120 gF: 0 vF: 0	) 37.0		317.5	241.5083		
StAOS1 SNP	681 Phytophthora infestans	2004 8.635			1	1		317.5	241.5083		
StAOS1 SNP StAOS2 SNP		-	7700297891	4E-4	(120/120 gF: 0 ∨F: 0	) 37.0					
StAOS1 SNP StAOS2 SNP StAOS2 SNP	678 Phytophthora infestans	2004 8.6357	7700297891 2135110180		(120/120 gF: 0 VF: 0) (120/120 gF: 0 vF: 0)			0.64	7.0E-4		
StAOS1 SNP StAOS2 SNP StAOS2 SNP StAOS2 SNP	678 Phytophthora infestans 774 MCR_2005	2004 8.6357 8.4532		61E-4	-	) -1.05	ò				
StAOS1 SNP StAOS2 SNP StAOS2 SNP StAOS2 SNP	678 Phytophthora infestans 774 MCR_2005 3 Globodera pallida cysts	2004 8.635 8.4532 2004 8.2048	2135110180	061E-4 687E-4	(120/120 gF: 0 vF: 0	) -1.05 ) 0.0	5	0.64	7.0E-4		

Figure 4: ConquestExplorer – Data analysis

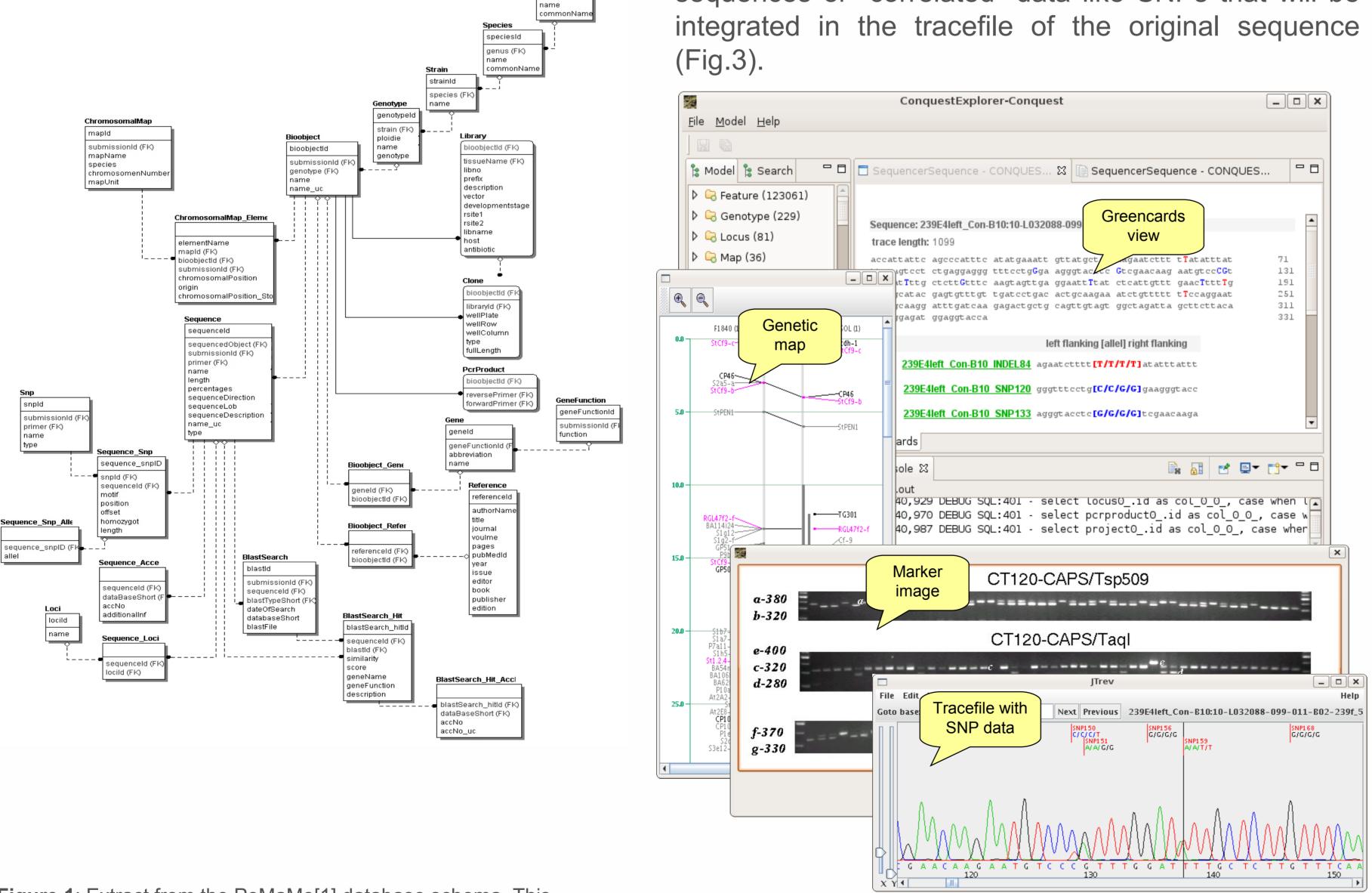
### Technologies

For the implementation of the ConquestExplorer we used mature and cutting-edge information technologies like

- an application platform (Eclipse Rich Client Platform),
- a database management system (HSQLDB),
- an object-oriented database interface (Hibernate),
- a fulltext search engine (Lucene) as well as
- analyzing software (GNU R).

#### Future Perspectives

ConquestExplorer will be extended to handle the new types of data generated within the Papatomics project like transcriptomic, proteomic and metabolomic data. We will also develop new or integrate available visualization tools for proteomic data. An interface to MapMan, a tool for visualization and pathway mapping of transcriptomic and metabolomic data, will be created. Furthermore, analysis tools for omics data will be developed and integrated.



**Figure 1**: Extract from the PoMaMo[1] database schema. This schema was modified for the ConquestExplorer database schema.



**Figure 3**: ConquestExplorer – Data visualization

#### References

[1] S. Meyer, A. Nagel and C. Gebhardt (2005) PoMaMo- a comprehensive database for potato genome data. Nucleic Acids Research 33 (Database Issue): D666-D670 [2] J. Neigenfind, G. Gyetvai, R. Basekow, S. Diehl, U. Achenbach, C. Gebhardt, J. Selbig and B. Kersten (submitted 2008) Haplotype inference from unphased SNP data in heterozygous polyploids based on the SAT. BMC Genomics, 9:356

[3] J. Neigenfind (2007) A generalized approach for calculating haplotypes in polyploid species based on the SAT-algorithm. Master Thesis, Free University Berlin