Implementation of a new data enrichment analysis tool in the Animal QTL Database

https://www.animalgenome.org/QTLdb

Zhi-Liang Hu, Carissa A. Park, and James M. Reecy

Department of Animal Science, Iowa State University, 2255 Kildee Hall, 806 Stange Road, Ames, IA 50011

Abstract

The development of Animal QTLdb over the past decade has provided a valuable resource for researchers to utilize a wealth of quantitative trait locus (QTL) and SNP association data to help elucidate potential genetic mechanisms that underlie traits of interest. The increasing growth of reported livestock animal QTL/association data in recent years provides new opportunities for data mining, meta-analysis, and big data analysis. We report here a new data enrichment analysis tool recently implemented in the QTLdb, which allows users to identify regions of the genome and to determine if they may be overrepresented in reported QTL/associations for traits of interest, similar to GO-enrichment testing of gene expression data. For example an analysis of 3,827 "milk yield" QTL/associations representing seven related traits found on 30 cattle chromosomes are "enriched" on chromosome 14. While the initial implementation of the enrichment tool demonstrates its utility, there are additional opportunities whereby this utility can be expanded by including more factors/parameters for more complex networked information analysis.

Background

Gene Ontology (GO) [1, 2] enrichment analysis has been used on large gene expression datasets to effectively identify implicated biological processes underlying experimental outcomes [3]. The method has been useful and well described [4]. In simplified terms, both a gene and a QTL represent a region of the genome. Likewise, Gene Ontology terms can be associated with genes just as phenotypes are associated with QTL. Therefore, we hypothesized that it would be possible to analyze genome regions for over-enrichment of a particular phenotype/trait.

Developments

As a preliminary trial, we assessed simple procedures to evaluate the enrichment of QTL/association data curated into the QTLdb using Chi-square analysis of a two-way contingency table (traits by chromosomes). The currently implemented tool was designed to allow evaluation of all reported QTL/associations for selected traits throughout a genome, to determine if the trait or traits are overrepresented in one or more regions. The analysis relies on an underlying assumption that the traits of interest are related. For example, they may belong to the same overarching trait type, or may be from a given trait ontology branch.

Figure 1 (a) shows the output from our initial implementation, in which 7,686 "growth" QTL/associations, representing 47 related traits (Figure 1 (b)) found on 30 cattle chromosomes, are "enriched" in certain chromosomal locations. Chi-square analysis was used to estimate the overrepresentation of reported QTL/associations in terms of their frequencies by chromosome locations. The contingency p-value (p) estimates are used to indicate the degree of overrepresentation (enrichment) of QTL/associations. The Benjamini-Hochberg procedure [5] is used to estimate the false discovery rate (FDR). The sizes of Chi-squares in each contingency category are graphically plotted with horizontal bars of varying lengths to indicate chromosomes or chromosome locations where larger numbers of QTL/associations are found.

If users are interested in sub-regions of a particular chromosome, they have the option to select a chromosome and give a window size in terms of the length of the chromosomal region to be evaluated (Figure 1 (a)).

Efforts are ongoing to expand functionality to allow options for users to select a set of traits for analysis based on their ontology reasoning. When available, trait correlation data are also appended to the enrichment results (Figure 1 (c)) to provide supporting information for user evaluation. This provides evidence to show that new potentials exist for further enrichment analysis involving factors that may be networked or related.

Access to the tool

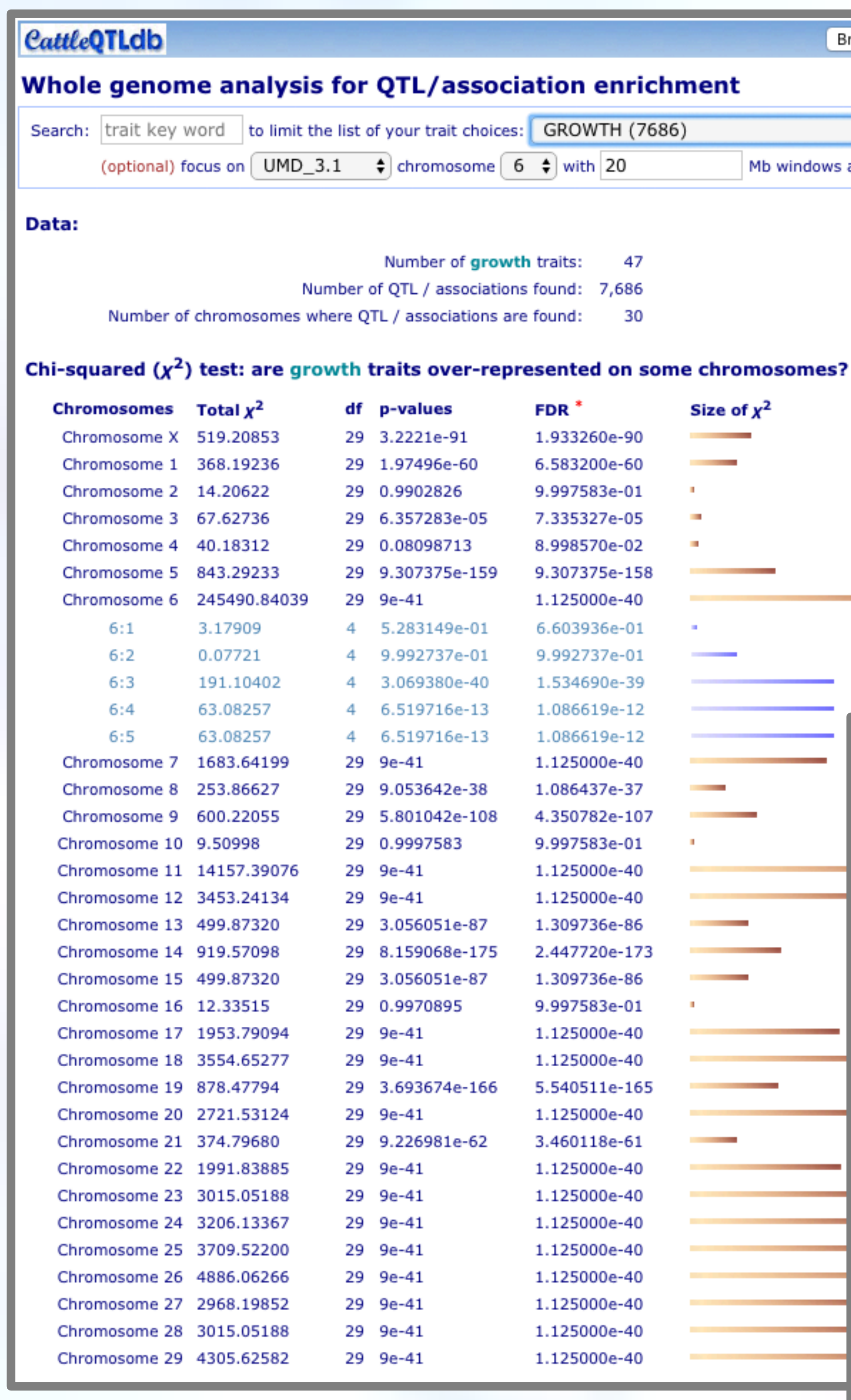
The data enrichment tool can be accessed via multiple paths on the Animal QTLdb website. The direct access to the tool is the "Whole genome enrichment analysis of QTL/associations" link in the "Search and Analysis" menu from a QTLdb species home page. There are also access points on other QTLdb data analysis tools where data that are currently being viewed can be further examined using the data enrichment analysis tool (refer to the website):

- "Search and Analysis" >> Search to "View genome distribution data for a trait" >> Click a trait type on the bottom list of the trait types >> "Try a genome wide enrichment analysis"
- "Search and Analysis" >> "Find related data by trait ontology search" >> Click a trait name for trait details view >> Explore Further: "Try a data enrichment analysis"
- "Data Browse": Open a trait hierarchy (Figure 2(b)) >> Navigate to a trait type of interest (Figure 2(a)) >> Click on a link-out icon >> "Find all QTL /association for this trait type" >> Below the genome plot there is a link to "Try enrichment analysis on this trait type" which will use the displayed trait data to start an enrichment analysis.

This is only a highlight of access routes. Other web access points to the enrichment analysis tool will also be created as the web site is continually developed.

Acknowledgments:

This work was supported in part by USDA Grant 2013-67015-21210 and USDA NRSP-8 Bioinformatics Coordination project.



CattleQTLdb

Browse | Search | View Maps | F A Q

Whole genome analysis for QTL/association enrichment

Search: trait key word to limit the list of your trait choices: GROWTH (7686) Submit

(optional) focus on UMD_3.1 chromosome 6 with 20 Mb windows across its length

Data:

Number of growth traits:	47
Number of QTL / associations found:	7,686
Number of chromosomes where QTL / associations are found:	30

Chi-squared (χ^2) test: are growth traits over-represented on some chromosomes?

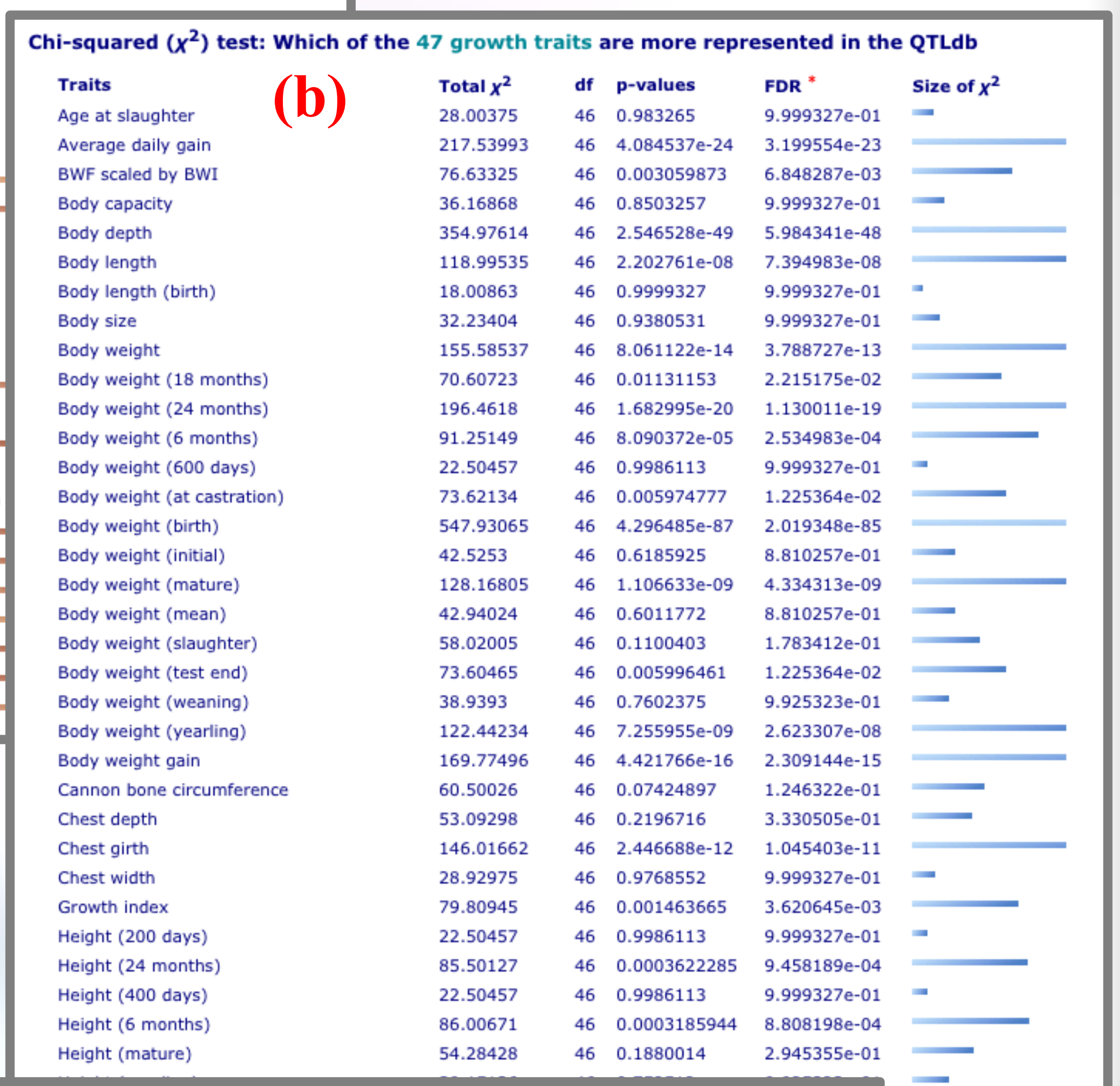
Chromosomes	Total χ^2	df	p-values	FDR *	Size of χ^2
Chromosome X	519.20853	29	3.2221e-91	1.933260e-90	
Chromosome 1	368.19236	29	1.97496e-60	6.583200e-60	
Chromosome 2	14.20622	29	0.9902826	9.997583e-01	
Chromosome 3	67.62736	29	6.357283e-05	7.335327e-05	
Chromosome 4	40.18312	29	0.08098713	8.998570e-02	
Chromosome 5	843.29233	29	9.307375e-159	9.307375e-158	
Chromosome 6	245490.84039	29	9e-41	1.125000e-40	
6:1	3.17909	4	5.283149e-01	6.603936e-01	
6:2	0.07721	4	9.992737e-01	9.992737e-01	
6:3	191.10402	4	3.069380e-40	1.534690e-39	
6:4	63.08257	4	6.519716e-13	1.086619e-12	
6:5	63.08257	4	6.519716e-13	1.086619e-12	
Chromosome 7	1683.64199	29	9e-41	1.125000e-40	
Chromosome 8	253.86627	29	9.053642e-38	1.086437e-37	
Chromosome 9	600.22055	29	5.801042e-108	4.350782e-107	
Chromosome 10	9.50998	29	0.9997583	9.997583e-01	
Chromosome 11	14157.39076	29	9e-41	1.125000e-40	
Chromosome 12	3453.24134	29	9e-41	1.125000e-40	
Chromosome 13	499.87320	29	3.056051e-87	1.309736e-86	
Chromosome 14	919.57098	29	8.159068e-175	2.447720e-173	
Chromosome 15	499.87320	29	3.056051e-87	1.309736e-86	
Chromosome 16	12.33515	29	0.9970895	9.997583e-01	
Chromosome 17	1953.79094	29	9e-41	1.125000e-40	
Chromosome 18	3554.65277	29	9e-41	1.125000e-40	
Chromosome 19	878.47794	29	3.693674e-166	5.540511e-165	
Chromosome 20	2721.53124	29	9e-41	1.125000e-40	
Chromosome 21	374.79680	29	9.226981e-62	3.460118e-61	
Chromosome 22	1991.83885	29	9e-41	1.125000e-40	
Chromosome 23	3015.05188	29	9e-41	1.125000e-40	
Chromosome 24	3206.13367	29	9e-41	1.125000e-40	
Chromosome 25	3709.52200	29	9e-41	1.125000e-40	
Chromosome 26	4886.06266	29	9e-41	1.125000e-40	
Chromosome 27	2968.19852	29	9e-41	1.125000e-40	
Chromosome 28	3015.05188	29	9e-41	1.125000e-40	
Chromosome 29	4305.62582	29	9e-41	1.125000e-40	

Figure 1 (a, left; and b, below)

Screenshot of enrichment analysis on "growth" QTL/association data. Note that the screenshots of (a), (b), and (c) are all on one web page.

In (a) chromosome 6 was selected for detailed enrichment analysis with a 20-Mbp window.

In (b) is shown how each of the 47 "growth" traits contributes to the genome location-wise enrichment analysis outcome.

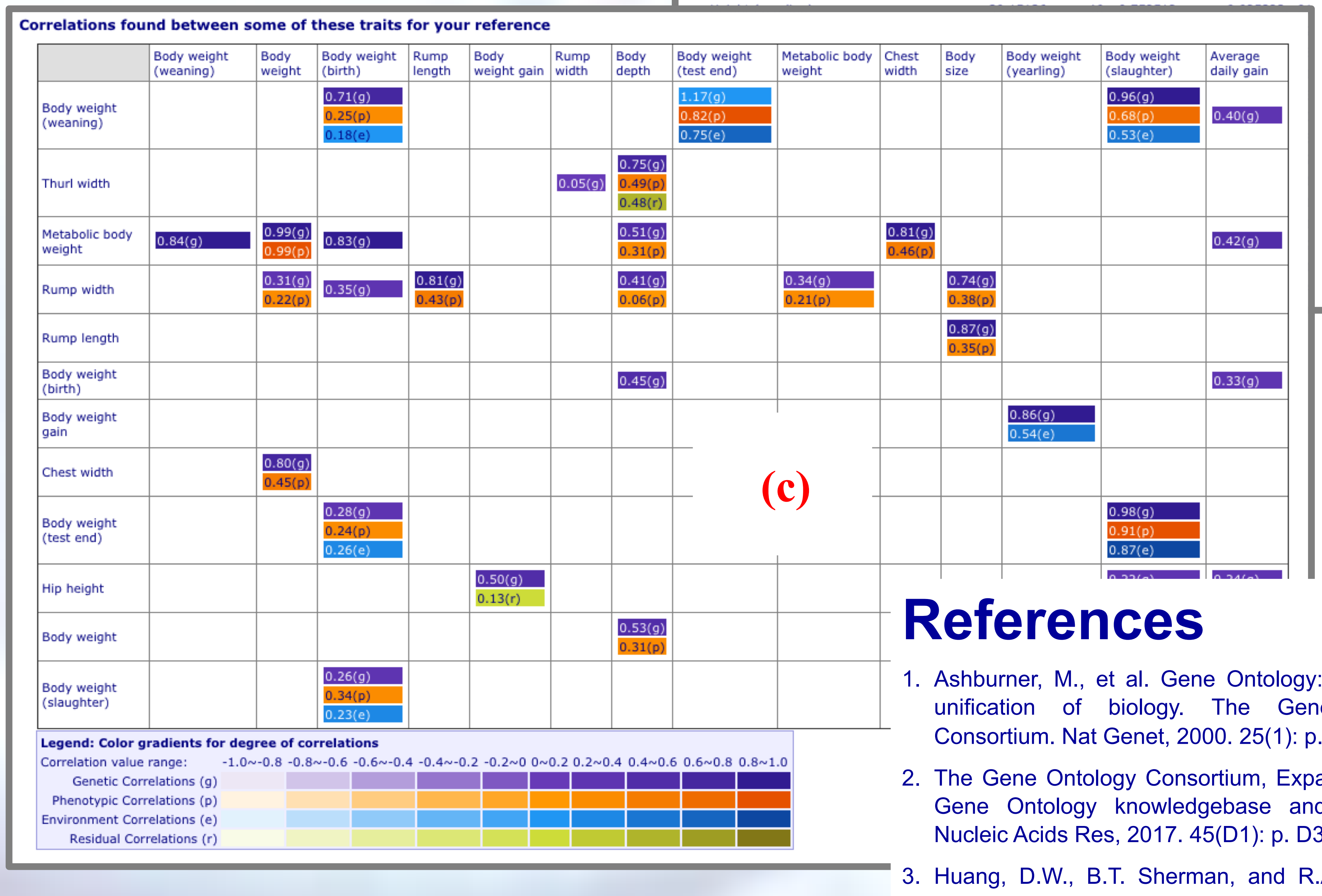


Chi-squared (χ^2) test: Which of the 47 growth traits are more represented in the QTLdb

Traits	Total χ^2	df	p-values	FDR *	Size of χ^2
Age at slaughter	28.00375	46	0.983265	9.999327e-01	
Average daily gain	217.53993	46	4.084537e-24	3.199554e-23	
BWF scaled by BWI	76.63325	46	0.003059873	6.848287e-03	
Body capacity	36.16868	46	0.8503257	9.999327e-01	
Body depth	354.97614	46	2.546528e-49	5.984341e-48	
Body length	118.99535	46	2.202761e-08	7.394983e-08	
Body length (birth)	18.00863	46	0.9999327	9.999327e-01	
Body size	32.23404	46	0.9380531	9.999327e-01	
Body weight	155.58537	46	8.061122e-14	3.788727e-13	
Body weight (18 months)	70.60723	46	0.01131153	2.215175e-02	
Body weight (24 months)	196.4618	46	1.682995e-20	1.130011e-19	
Body weight (6 months)	91.25149	46	8.090372e-05	2.534983e-04	
Body weight (600 days)	22.50457	46	0.9986113	9.999327e-01	
Body weight (at castration)	73.62134	46	0.005974777	1.225364e-02	
Body weight (birth)	547.93065	46	4.296485e-87	2.019348e-85	
Body weight (initial)	42.5253	46	0.6185925	8.810257e-01	
Body weight (mature)	128.16805	46	1.106633e-09	4.334313e-09	
Body weight (mean)	42.94024	46	0.6011772	8.810257e-01	
Body weight (slaughter)	58.02005	46	0.1100403	1.783412e-01	
Body weight (test end)	73.60465	46	0.005996461	1.225364e-02	
Body weight (weaning)	38.9393	46	0.7602375	9.925323e-01	
Body weight (yearling)	122.44234	46	7.255955e-09	2.623307e-08	
Body weight gain	169.77496	46	4.421766e-16	2.309144e-15	
Cannon bone circumference	60.50026	46	0.07424897	1.246322e-01	
Chest depth	53.09298	46	0.2196716	3.330505e-01	
Chest girth	146.01662	46	2.446688e-12	1.045403e-11	
Chest width	28.92975	46	0.9768552	9.999327e-01	
Growth index	79.80945	46	0.001463665	3.620645e-03	
Height (200 days)	22.50457	46	0.9986113	9.999327e-01	
Height (24 months)	85.50127	46	0.0003622285	9.458189e-04	
Height (400 days)	22.50457	46	0.9986113	9.999327e-01	
Height (6 months)	86.00671	46	0.0003185944	8.808198e-04	
Height (mature)	54.28428	46	0.1880014	2.945355e-01	

Figure 1 (c, below)

Correlation data for the displayed traits is included when available.



Correlations found between some of these traits for your reference

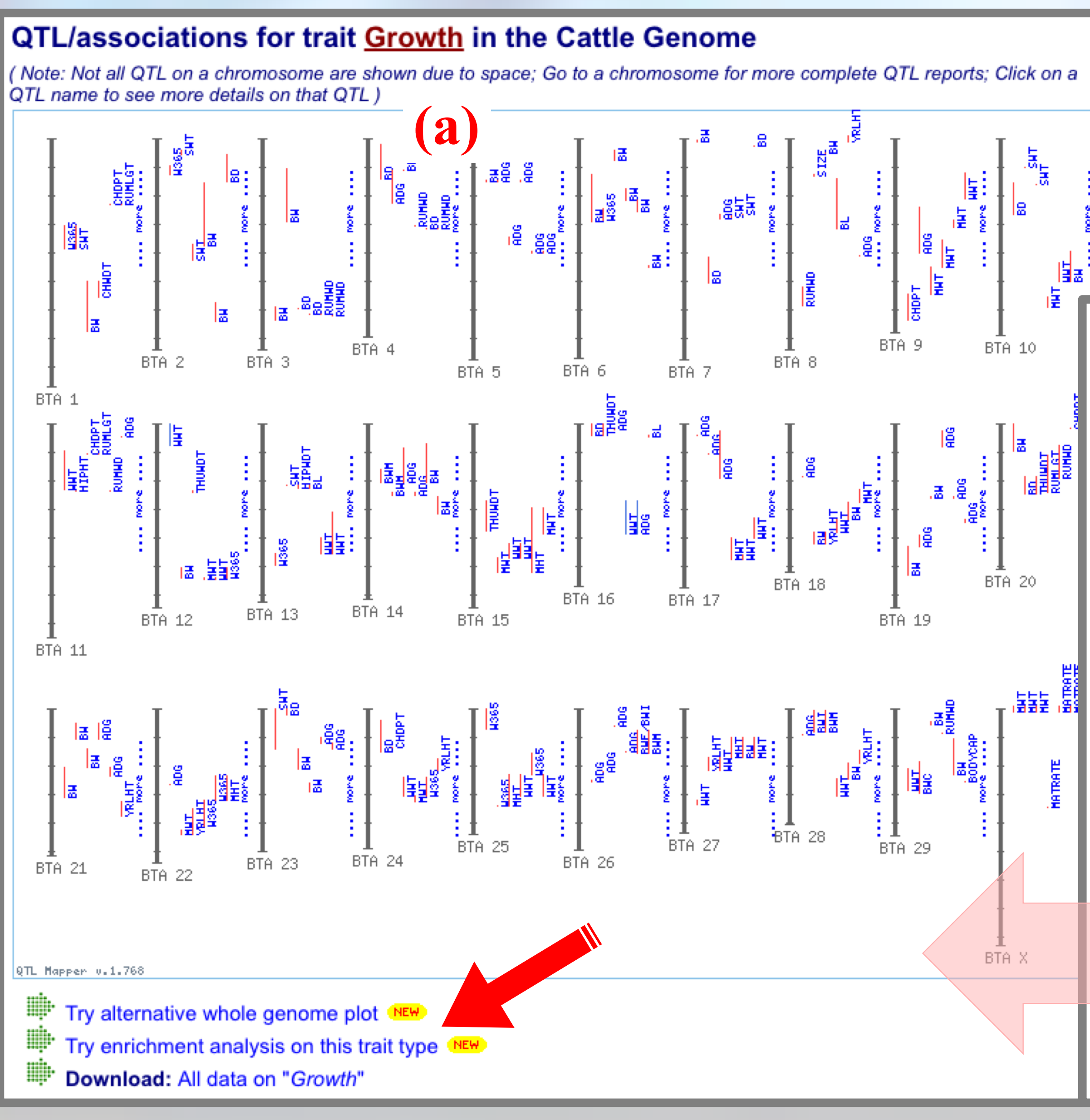
	Body weight (weaning)	Body weight	Body weight (birth)	Rump length	Body weight gain	Rump width	Body depth	Body weight (test end)	Metabolic body weight	Chest width	Body size	Body weight (yearling)	Body weight (slaughter)	Average daily gain
Body weight (weaning)			0.71(g) 0.25(p) 0.18(e)					1.17(g) 0.82(p) 0.75(e)					0.96(g) 0.68(p) 0.53(e)	0.40(g)
Thurl width						0.05(g)	0.75(g) 0.49(p) 0.48(r)							
Metabolic body weight	0.84(g)	0.99(g) 0.99(p)	0.83(g)				0.51(g) 0.31(p)			0.81(g) 0.46(p)				0.42(g)
Rump width		0.31(g) 0.22(p)	0.35(g)	0.81(g) 0.43(p)			0.41(g) 0.06(p)		0.34(g) 0.21(p)		0.74(g) 0.38(p)			
Rump length											0.87(g) 0.35(p)			
Body weight (birth)							0.45(g)							0.33(g)
Body weight gain												0.86(g) 0.54(e)		
Chest width		0.80(g) 0.45(p)												
Body weight (test end)			0.28(g) 0.24(p) 0.26(e)										0.98(g) 0.91(p) 0.87(e)	
Hip height					0.50(g) 0.13(r)								[illegible]	[illegible]
Body weight							0.53(g) 0.31(p)							
Body weight (slaughter)			0.26(g) 0.34(p) 0.23(e)											

Legend: Color gradients for degree of correlations

Correlation value range:	-1.0~-0.8	-0.8~-0.6	-0.6~-0.4	-0.4~-0.2	-0.2~0	0~0.2	0.2~0.4	0.4~0.6	0.6~0.8	0.8~1.0
Genetic Correlations (g)										
Phenotypic Correlations (p)										
Environment Correlations (e)										
Residual Correlations (r)										

Figure 2

Links to the enrichment tool from other tools: from genome-wide data plot (a, below), and from the trait ontology browser (b, lower right).



QTL/associations for trait Growth in the Cattle Genome

(Note: Not all QTL on a chromosome are shown due to space; Go to a chromosome for more complete QTL reports; Click on a QTL name to see more details on that QTL)

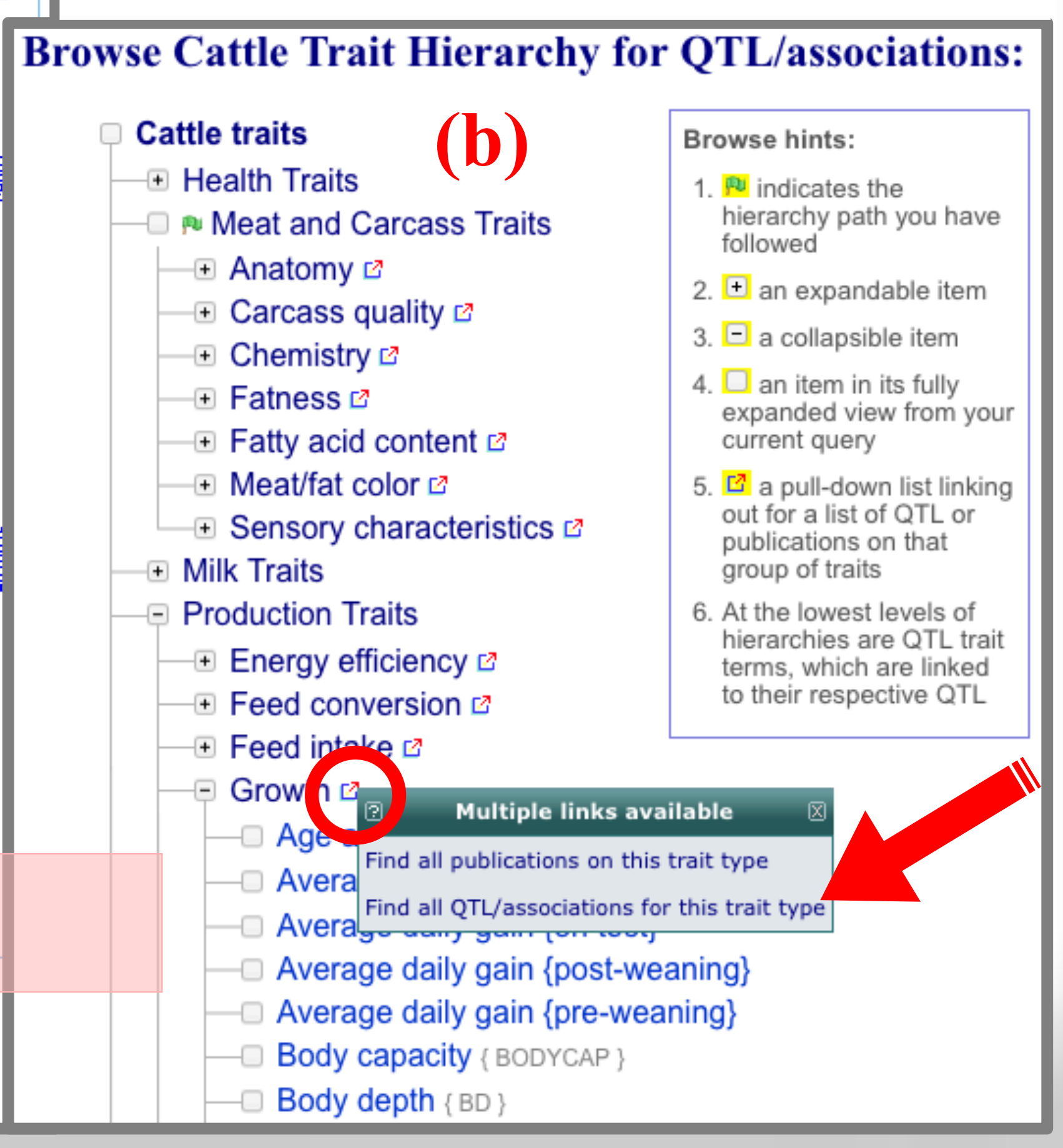
Try alternative whole genome plot NEW

Try enrichment analysis on this trait type NEW

Download: All data on "*Growth*"

References

1. Ashburner, M., et al. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet, 2000. 25(1): p. 25-9.
2. The Gene Ontology Consortium, Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res, 2017. 45(D1): p. D331-D338.
3. Huang, D.W., B.T. Sherman, and R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res, 2009. 37(1): p. 1-13.
4. Subramanian, A., et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA, 2005. 102(43): p. 15545-50.
5. Benjamini, Y. and Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol, 1995. 57(1).



Browse Cattle Trait Hierarchy for QTL/associations:

- Cattle traits
  - ⊞ Health Traits
  - Meat and Carcass Traits
    - ⊞ Anatomy
    - ⊞ Carcass quality
    - ⊞ Chemistry
    - ⊞ Fatness
    - ⊞ Fatty acid content
    - ⊞ Meat/fat color
    - ⊞ Sensory characteristics
  - ⊞ Milk Traits
  - ⊟ Production Traits
    - ⊞ Energy efficiency
    - ⊞ Feed conversion
    - ⊞ Feed intake
    - ⊟ Growth
      - Age
      - Avera
      - Avera
      - Average daily gain {post-weaning}
      - Average daily gain {pre-weaning}
      - Body capacity { BODYCAP }
      - Body depth { BD }

Multiple links available

Find all publications on this trait type

Find all QTL/associations for this trait type

Browse hints:

1. indicates the hierarchy path you have followed
2. ⊞ an expandable item
3. ⊟ a collapsible item
4. □ an item in its fully expanded view from your current query
5. a pull-down list linking out for a list of QTL or publications on that group of traits
6. At the lowest levels of hierarchies are QTL trait terms, which are linked to their respective QTL