International equine gene mapping workshop report: a comprehensive linkage map constructed with data from new markers and by merging four mapping resources


International equine gene mapping workshop report: a comprehensive linkage map constructed with data from new markers and by merging four mapping resources

Manuscript received 26 August 2004; accepted in revised form for publication by M. Schmid 26 August 2004.

Abstract. A comprehensive male linkage map was generated by adding 359 new, informative microsatellites to the International Equine Gene Map half-sibling reference families and by combining genotype data from three independent mapping resources: a full sibling family created at the Animal Health Trust in Newmarket, United Kingdom, eight half-sibling families from Sweden and two half-sibling families from the University of California, Davis. Because the combined data were derived primarily from half-sibling families, only autosomal markers were analyzed. The map was constructed from a total of 766 markers distributed on the 31 equine chromosomes. It has a higher marker density than that of previously reported maps, with 626 markers linearly ordered and 140 other markers assigned to a chromosomal region. Fifty-nine markers (7%) failed to meet the criteria for statistical evidence of linkage and remain unassigned. The map spans 3,740 cM with an average distance of 6.3 cM between markers. Fifty-five percent of the intervals are ≤ 5 cM and only 3% ≥ 20 cM. The present map demonstrates the cohesiveness of the different data sets and provides a single resource for genome scan analyses and integration with the radiation hybrid map.
Horse gene mapping research, organized as the International Equine Gene Mapping Workshop, has taken place under the auspices of The Dorothy Russell Havemeyer Foundation and the collaboration among scientists worldwide. The goal of the workshop has been to develop and make available to scientists basic resources such as genetic maps that are critical to genomic research. Horse genomic information thus developed has been used for applications to map phenotypic traits as well as to study the natural history of horses (Vila et al., 2001; Chowdhry and Bailey, 2003). One of the primary resources of the workshop is the horse linkage map.

Four linear maps for the horse genome have been published, including three linkage maps with 140 (Lindgren et al., 1998), 353 (Swinburne et al., 2000a) and 344 (Guérin et al., 1999, 2003) markers and a radiation hybrid (RH) map with 730 markers (Chowdhary et al., 2003). Cytogenetic and comparative mapping research has further contributed to the integrity of the linear genetic maps and extended the genomic information for the horse (reviewed in Chowdhary and Bailey, 2003). Comparison of the maps and use of the information has been facilitated by overlap of markers contained in each of the maps. In this report we describe the expansion of the workshop linkage map to 766 loci based on 839 markers derived from new and previously published linkage data to create a comprehensive map. The map constructed by this approach provides expanded coverage of horse chromosomes that will significantly improve the genomic information for the horse. The integration of all available linkage data into one map benefits scientists by facilitating the identification and selection of markers that are critical for discovery of genes associated with phenotypic traits, health and performance in the horse.

Materials and methods

Reference family panels

Twenty-four families with a total of 921 offspring were used to construct the linkage map. They consisted of the 13 paternal half-sib families with 500 offspring from the International Horse Reference Family Panel (IHRFP) (Guérin et al., 1999, 2003), the eight paternal half-sib reference families with 262 offspring from Uppsala, Sweden (SRF) (Lindgren et al., 1998), the two three-generation, full-sib reference families with 67 offspring created at the Animal Health Trust, Newmarket, UK (NRF) (Swinburne et al., 2000a) and two paternal half-sib Quarter Horse families with 92 offspring from the Veterinary Genetics Laboratory, University of California, Davis (VGL) (Locke et al., 2002).

Markers and analysis

The genotypic data assembled contained 839 informative markers representing 825 autosomal and 14 X-linked loci. The set of markers was obtained by combining genotype files for 344 loci from the two workshop reports (Guérin et al., 1999, 2003), 144 markers from the SRF, 353 markers from the NRF, 100 markers from VGL and by genotyping 359 new microsatellites on the IHRFP resource of which 278 (77%) were typed by one laboratory (code DAV). A complete list of markers and source of data for this report is shown in supplement Table S1 available at http://www.uky.edu/Horsemap/ConsensusMap. For the purpose of merging and analyzing data, the genotype file was edited to assign a common name to markers representing the same locus but having different names in separate mapping resources. For example, HMB2, HM83, HMB4, HMB5, HMB6 and Ext in SRF data correspond to AHT002, AHT003, AHT004, AHT005, AHT006 and MCIr in other mapping resources, including cytogenetic and RH maps. For these cases, the latter names were used as locus identifiers in the SRF set. A similar approach was used for other loci and for all instances the locus name of choice followed the nomenclature in the Horsemap database (http://locus.jouy.inra.fr/cgi-bin/gbc/mapping/common/main.pl?BASE=horse).

In the course of testing, three pairs of DNA samples from the IHRFP, each from a different family, were found to have almost identical results (1001 and 10031, 13039 and 13045, 20008 and 20001). The most likely explanation is that three horses were accidentally sampled in duplicate. Therefore, individuals 10031, 13045 and 20008 were removed from the data set. Individual 10031 was excluded because X-linked markers showed it to be female (same sex as 10029) and not male as specified in the pedigree file. Individuals 13045 and 20008 were excluded because they had fewer reported genotypes and contributed fewer meioses than their duplicates, perhaps because of lower quality or quantity of the extracted DNA. A list of the new markers typed on the IHRFP is given in Table S2 as supplementary information available at http://www.uky.edu/Horsemap/ConsensusMap.

Relative to the total number of autosomal markers in the data set, 294 markers (36%) were typed in two or more mapping resources, 430 markers (52%) were typed only in the IHRFP, 80 markers (10%) were typed only in the NRF and 20 markers (2.4%) were typed only in the SRF. Dams were available for two IHRFP families, the NRF family and two VGL families and their genotypes were included in the analyses only to help identify paternally transmitted alleles. Because of the paternal half-sib structure of mapping resources, except for NRF, only the recombination fractions for autosomal markers obtained through male meioses were considered to construct the comprehensive map.

The CRIMAP program version 2.4 (Green et al., 1990) was used for linkage analysis. The program was modified to handle the increased number of loci (n > 99) and to adjust memory allocations. Maximum likelihood estimates of recombination fraction (theta) were calculated using the TWO-POINT option with a significant lod-score threshold >3 to determine the appropriate linkage group association and chromosome assignment of 505 markers not previously included in the Phase II framework map (Guérin et al., 2003). Multi-point analysis with the BUILD option was then used to insert new loci within each chromosome, starting with a lod-score threshold of >3. The best linear order was finally determined with a lod-score of >0.5 and checked with the FLIPS option. When the lowest negative log10_likelihood was obtained for each of the 31 autosomes, the BUILD option was rerun to check for additional insertions. This routine was repeated until the best map was produced and no new marker appeared in the linear order. The CHROMPIC option was used in a few instances to identify potential genotyping errors in the IHRFP data that could result in unlikely recombination events, specifically, multiple recombinants within a small region. In six of these cases, new genotype data were collected by retyping markers COR028, COR058, COR062, COR100, LEX033 and UM004 at the Veterinary Genetics Laboratory (code DAV). Map distances were calculated using the Kosambi function restricted to the male-specific map.

Map representation

The linkage map was constructed based on the best final linear order and distances. Cytogenetic information of physical assignments made by fluorescence in situ hybridization (FISH) was used to orient chromosomes. The cytogenetic information reviewed, compiled and expanded by Chowdhary et al. (2003) was used as the basic reference with additional information supplemented from other sources (http://locus.jouy.inra.fr/cgi-bin/gbc/mapping/common/intro2.pl?BASE=horse and http://www.thearkdb.org/).

Results and discussion

A total of 825 informative autosomal markers were analyzed, of which 766 (93%) were unambiguously assigned to one of 31 linkage groups. Of these markers, 626 (82%) were linearly ordered and 140 (18%) were assigned to a chromosomal region but could not be inserted into the linear map with sufficient statistical support. Fifty-nine markers (7%) failed to meet the criteria for statistical evidence of linkage and remain unassigned. The total number of informative meioses (IM) was 182,467 with average number of IM of 248.48 ± 135.16 for ordered loci, 166.69 ± 106.69 for assigned loci and 60.25 ±
49.88 for unassigned loci. Results from single factor ANOVA for differences in average number of IM (data available in supplement Table S3 at http://www.uky.edu/Horsemap/ConsensusMap) between the three categories of marker mapping status were highly significant ($F = 75.27, df 2, 822, P = 9.6 \times 10^{-13}$) as were those between ordered and assigned ($F = 44.76, df 1, 764, P = 4.29 \times 10^{-11}$). These analyses suggested that pooling the mapping resources and the concomitant increase in number of informative and co-informative meioses for common markers had a positive effect in the ability to insert and order loci in the map.

The linkage map spanning about 3,740 cM and oriented according to the cytogenetic map is shown in Fig. 1. While the number of markers in common between the two maps was sufficient to help orient most linkage groups, many chromosomes were poorly represented in the cytogenetic map (e.g. ECA5, 6, 7, 8, 9, 12, 13, 21, 24, 25, 28, 29, 30). An increase in the number of physically mapped markers, as well as more dispersed FISH assignments, would provide useful complimentary information for accurate alignment of linkage groups and validation of the order of loci obtained in this study.

With few exceptions, there was good agreement in linear order of markers with previously published linkage maps. The average distance between ordered markers is $6.3 \pm 5.8$ cM (range 0–31 cM) with 56% of the intervals $\leq 5$ cM and only 3% $\geq 20$ cM. Details of coverage and map length by chromosome are shown in Table 1. Relative to the previous workshop map (Guérin et al., 2003), marker density increased 2.4-fold but the percentage of ordered and assigned loci remained essentially the same at about 75 and 16%, respectively. Seventeen chromosomes (ECA1, 2, 3, 4, 8, 9, 10, 12, 14, 15, 16, 18, 20, 22, 23, 30 and 31) contain one to three regions with marker intervals between 18 and 31 cM with the largest gaps located on ECA3 between HTG002 and TKY780 (31 cM) and ECA8 between TKY436 and UCDEQ046 (30 cM). Identification of informative markers to reduce intervals and to improve coverage in the 17 chromosomes listed above should be an important consideration for future horse gene mapping research. This could be accomplished by targeted mapping of new polymorphic microsatellites on the HRRF for which chromosome assignments are known from two-point linkage analyses on the NRF resource (Mickelson et al., 2003; Swinburne et al., 2003) or RH panel (Wagner et al., 2004) or by targeted development of markers using high resolution RH maps to identify genes within the regions, select BACs containing those genes and search for microsatellites within those BACs.

With few exceptions, there was good agreement in linear order of markers with previously published linkage maps. The following differences are notable: VHL078 was ordered on ECA2 (Guérin et al., 2003) but for the current map it was only assigned to that chromosome as it could not be inserted in the linear map with significant odds. The likely position of VHL078 near ASB017 was suggested by results from TWO-POINT analyses. HTG020 mapped to ECA7 based on significant linkage to LEX038, LEX015 and TKY506, and not to ECA5 as previously reported (Swinburne et al., 2000a). NVHEQ067 mapped to ECA10 based on significant linkage to TKY496, TKY503 and TKY722, and not to ECA22 as previously reported (Guérin et al., 2003). Furthermore, the current assignment of NVHEQ067 agreed with the RH map assignment (Chowdhary et al., 2003).

Discrepancies between FISH and linkage map locations have been noted in published linkage maps for markers AHT030, AHT044, ASB014, ASB038, SGCV008, SGCV032, and TKY028. AHT030 was initially localized on ECA22 by FISH (Swinburne et al., 2000b) but previous workshop linkage data (Guérin et al., 2003) and this report (with increased number of meioses) placed this marker on ECA13. Incorrect FISH locations of AHT044, ASB014 and ASB038 attributed to use of chimeric clones as probes have been already been discussed (Swinburne et al., 2000a). Published linkage maps and this report supported location of these markers on ECA11, ECA8 and ECA27, respectively. Similarly, SGCV008 was physically mapped to ECA19 (Godard et al., 1997) but linkage data for this marker (Swinburne et al., 2000a; Guérin et al., 2003), including this report, supported its location on ECA10.

New FISH location of SGCV032 to ECA8 has been reported (Guérin et al., 2003) that is now in agreement with linkage data. The discrepancy for TKY028 has been noted (Guérin et al., 2003).
Fig. 1. A male comprehensive linkage map of horse autosomes. Orientation of linkage groups is based on FISH assignments shown on the cytogenetic map displayed next to G-banded schematic drawings of individual chromosomes on left (vertical lines depict physical location). To optimize display, the anchor loci depicted are a subset of the markers in common with the linkage map. Display of all common loci is available at http://www.vgl.ucdavis.edu/equine/caballus/. Idiogram nomenclature is taken from the standardized karyotype (Bowling et al., 1997). Gray solid lines connect markers on both maps. Ordered markers are connected to the bar diagram by a solid line. Markers not separated by recombination are shown with a branched marker position. Vertical solid lines to the right of each linkage group indicate likely location of assigned loci. New markers added to the map are shown in black font.

and linkage data available for this marker supported its location on ECA6 and not ECA10 as suggested by FISH results (Kakoi et al., 2000). Our assignment of HMS076 to ECA14, supported by significant linkage to COR103 (theta = 0, lod = 3.61) and COR104 (theta = 0, lod = 7.83), indicated another potential discrepancy since this marker was located on ECA5 by FISH (Mariat et al., 2001). However, we view our linkage assignment as tentative because of the low number of informative meioses (46) for HMS076 contributed by a single family. Improvement of genotype data by testing this marker on another resource family may help pinpoint its correct linkage assignment.
The total length of 3,740 cM for the map exceeded projections of about 2,700 cM suggested by Lindgren et al. (1998) and stands as one of the longest male-based maps among livestock species. A longer male map with 1,015 markers and spanning 3,876 cM has been reported in sheep (Maddox et al., 2001). However, comparisons among maps are difficult since recombination rates may vary among (Pardo-Manuel de Villena and Sapienza, 2001) and within (Lien et al., 2000; Weiman et al., 2003) species. Furthermore, different strategies have been used to build mapping resources and the number of individuals contributing to map development also varied. The 24 families used to construct the present map represent several different breeds of horses. Genotyping errors that inflate recombination fractions between markers are a factor for map length expansion. In
humans, it has been estimated that ~0.08% of genotype errors, primarily caused by mistyping but also by mutation or gene conversion, could increase map length by ~25% (Broman et al., 1998). Comparison of genotype data for duplicate samples in the IHRFP suggested an error rate of about 0.6% affecting male recombination. Although we identified and corrected some errors that resulted in unlikely recombination events, we recognize that genotyping errors still present in the data set might have artificially expanded the map length. Nevertheless, the order of markers obtained for each chromosome is consistent with data from cytogenetic and RH maps.

Progress in the development of linkage maps for other live-stock species, such as pig, cattle and sheep, has been achieved through merger of data from independent mapping resources (e.g. Kapke et al., 1996; Campbell et al., 2001; Casas et al., 2001; Kurar et al., 2002 to name only a few examples) or by combining genotype data for one mapping resource collected by different laboratories (e.g. Maddox et al., 2001). These mapping efforts share the objective of integrating genetic information and developing, as a community resource, high-resolution linkage maps that allow more efficient mapping of single gene and polygenic traits. The goals and efforts of the present horse gene mapping workshop to combine independent mapping data and to increase resolution of the linkage map parallel those of other species.

The linkage map described herein, albeit still at low density, represents a significant improvement over previous maps and provides a more useful resource from which to select markers for genomic analyses to map traits in the horse. Most of the new markers added to the IHRFP have also been recently typed on the RH panel and, as a new generation of the RH map becomes available, integration of the two maps will be possible to obtain a comprehensive map with physically ordered markers. The increase in map density will allow for better resolution of marker order and distance, particularly in regions where distances between adjacent markers are greater than 18 cM. Future workshop activities should address these issues.

Acknowledgements

We most gratefully acknowledge Irmina Bienkowska, Leah Brault, Stephanie Bricker, Ewa Iwanczyk, Janelle Katz, Liv Midthjell, Rosane Oliveira, Dianna Pettigrew, Elaine Philbrick, Angel del Valle, Thea Ward and Amy Young for their assistance with marker development and genotyping. We also thank Aaron Wong for his help with illustrations.

References


Chowdhary BP, Bailey E: Equine genomics: galloping.


Greens F, Fall KA, Crooks S: Documentation for CRI-MAP, Version 2.4. (Washington University School of Medicine, St. Louis 1990).


Greens F, Fall KA, Crooks S: Documentation for CRI-MAP, Version 2.4. (Washington University School of Medicine, St. Louis 1990).