Putting a Dent in Our Understanding of Maize Kernel Morphology

Honors Thesis

Presented to the College of Agriculture and Life Sciences, Plant Sciences Department

Of Cornell University

In Partial Fulfillment of the Requirements for the

Research Honors Program

By:

Matthew D. Murray

May 2013

Edward S. Buckler

Abstract

Much of our modern maize germplasm was originally brought about by the combination of northern flint lines and southern dent lines. Yet commercial production in the US today is dominated by dent or semi-dent kernel type maize (Corn Belt dent), which has hard outer walls of endosperm surrounding a soft floury interior that when dried compacts to form the characteristic dent in the top of the kernel. One major exception is flint type corn, which is grown in areas of North America, Europe, South America, the Caribbean and many parts of Africa. Flint maize is characterized by its rounded vitreous outer endosperm and soft granular center and has desirable qualities such as cold tolerance, disease and insect resistance as well as longer storage capacity than many dent lines.

The Nested Association Mapping (NAM) population parent inbred lines represent many of the major kernel types found in maize. In 2006 the entire NAM population of recombinant inbred lines were grown and scored visually for kernel type, in five locations. The NAM population contains 25 parents, of which, there are nine flint, nine semi-dent, four dent, two sweet and one popcorn parent lines, with a common dent parent, B73. Linkage mapping with ~7400 intervals markers was used to examine the genetics of kernel morphology. This yielded several areas of the genome that are significantly associated with the difference in kernel type seen in NAM. Several other major and minor QTL are shared across many families. Genome wide association study (GWAS) was also used in the maize association panel. Suggested peaks from both linkage mapping and GWAS highlight 13 candidate genes in starch and protein related pathways in the endosperm.

Introduction

Maize kernel morphology is one of the oldest distinguishing characteristics of maize. Many of our traditional races of maize are even differentiated and named by their kernel morphology. Kernel morphology was also one of the first characteristics studied in maize [1], and *sugary1* (*su1*), an endosperm mutant that changes the kernel morphology drastically, was one of first genes to be characterized genetically [2]. Difference in maize kernel morphology is often very easy to distinguish, as many of the mutants cause large physical changes in texture and appearance to the kernels. Some of the major mutants are the sugary, opaque, waxy, and floury families of genes. Although these mutants were some of the first traits to be studied, the genetics and biochemistry underlying these mutants are still under active investigation today [3], [4], [5].

Kernel morphology mutants are easy to spot visually, which led to early studies of these traits. For example, *su1*, the mutaion that gave rise the original sweet corns, and that leaves kernels shriveled upon desiccation, was first observed to have genetic properties over a century ago by Hays [1]. Another mutant, *waxy*, which gives the kernels a waxy texture, was brought to the US in the early 1900's from China [6], and its properties of inheritance were documented soon after [7]. Similarly *floury1(fl1)* was first reported early on [8], and *floury2(fl2)* was reported many years later [9], both for their characteristic floury endosperm. The *opaque2(o2)* mutant was described by the same group around the same time [10]. These mutants all lead to severe kernel morphology changes, and are visible to the

human eye, before and after harvest. These have made for great studies throughout the history of maize genetics and are still model systems recognized today.

The biochemistry of maize kernels also has a long history. The kernel mutants discussed above lend themselves to biochemical studies. These studies have been done since chemical fractional methodologies have been able to discern individual components of maize endosperm [11]. The endosperm in a maize kernel makes up 80-85% of the dry weight [12]. Of that 80-85%, 70-90% is starch [13]. After starch the next most abundant category is storage proteins, at 5-10% in non-mutant lines [13].

There are two forms of starch in maize, amylose and amylopectin. Starch is a polymeric molecule made up of glucose monomers. When these monomers join to form α-(1,4) carbon linkages, they make straight chain polymers [14]. When α-(1,6) carbon linkages are formed, a branched molecule results. Amylose and amylopectin have both α-(1,4) and α-(1,6) linkages, but in amylose there is less than 1% branching α-(1,6) bonds [14]. Generally amylose is only 100-10,000 monomers long, where amylopectin is 10,000 to 100,000 monomers per polymer [15]. The *su1* and *waxy* mutants result from large differences in starch biosynthesis and storage.

The second largest fraction of the maize endosperm consists of the storage proteins in maize. In maize the most common storage proteins are the zeins [16]. There are four classes [16], and a wide range in size of the storage proteins [16], [17], [18]. The physical structures of different zeins have been studied [19], [18]. Zeins are synthesized in the endoplasmic reticulum [20] and as kernels begin to dry and cells in the endosperm die, the proteins are left in the endosperm and can form complexes with starches [18]. Protein differences are responsible for many mutants such as the *o2* and *fl1*, *fl2* and *fl3*.

Modern commercial maize in the US can be classified as Corn Belt Dent. Corn Belt Dent varieties are derived from the combination of two ancestral prominent races of maize in the US, Northern Flint and Southern Dent [21]. The dent character was chosen by selection in the US over the last century, while the flint type was selected elsewhere in the world, and is still prominent in regions such as Argentina, parts of Africa, the Caribbean and most of Europe [22].

Dent kernels, when dry, have a small indentation on the crown of the kernel. In dent kernels, the outer walls of the endosperm are hard and corneous, while the interior of the kernel is made up of soft, floury endosperm. When this soft, floury endosperm dries, the loss of water causes the endosperm to compact, which allows the center of the kernel to shrink and become indented [22]. The degree to which a kernel is dented varies based on the genetic background. Flint kernels are similar to dent kernels in the side walls of the kernel, but differ in the central and top of the kernel endosperm. Flint kernels are characterized by a hard vitreous endosperm in the upper surface of the kernel with a soft granular starch center endosperm [22].

The biological differences between flint and dent kernels have been well documented. Dent kernels are slightly more efficient for processing and provide a higher digestibility index to livestock [23]. Flint kernels were prominent in the colder, northeastern parts of the US, and can give some germination

advantage in colder climates [22], [24]. Flint kernel types also been found to have some extra natural insect resistance compared to dent kernel types [25].

Maize kernel biochemistry and kernel mutants have been well characterized and studied for more than a century, yet the quantitative genetics controlling the natural variation accounting for the differences in kernel morphology are poorly understood. The NAM (Nested Association Mapping) population is a large, diverse and statistically powerful mapping population [26], [27], [28] that encompasses dent, flint and intermediate semi-dent kernel morphologies. This population has been used to map many important traits in maize [27], [29], [30], [31]. The Maize Association Panel [32] is a diverse population from which NAM was derived that has also been used to map many important traits in maize [29], [30], [31], [33]. A wealth of sequence data has been published on these populations ([34], [35], [36] http://www.panzea.org/, http://maizesequence.org/index.html). Several statistical methodologies have also been developed to map and explain the genetic contributions of diverse phenotypes [26], [27]. When these diverse and powerful populations are evaluated for kernel morphology, and the current genomic and statistical approaches are applied to this phenotype, it will be possible to obtain a better understand of the natural genetic variation of this trait. This study will use current statistical methods to evaluate the phenotypic values of two diverse maize populations and locate candidate genes controlling the genetic variation underlying this phenotype.

Materials and Methods

Germplasm

The Maize Association Panel consists of 282 diverse inbred maize lines. This population has previously been described [32]. This population has a wide variety of kernel types including dent, semi-dent, flint, sweet, and popcorn type kernels. The NAM population was developed from a 25 line subset of the association panel. The NAM population consists of 5,00 Recombinant Inbred Lines (RILs) developed by crossing the 25 founder lines to B73, followed by five generations of selfing. This population has been previously described [28]. The NAM founders have at least a single representative from each of the five kernel types mentioned in the Association Panel.

Phenotypic Data

The Association Panel and NAM population were grown in five environments in 2006 and 2007. Field locations consisted of Aurora, NY (2006, summer), Clayton, NC (2006, summer), Homestead, FL (2007, winter), Urbana, IL (2006, summer), and Ponce, PR (2007, winter). Kernel type was recorded for each entry. Kernels were observed and recorded into one of the following categories: "Dent," "Flint," "Semi-Dent," "Sweet," or "Popcorn." For the purpose of this study, sweet and popcorn lines were removed. This removes three of the NAM populations: B73xHP301 popcorn, B73xIl14h and B73xP39 sweet types. These families' kernel types are outside of the immediate scope of this study, and therefore, these families were left out of the mapping exercises. For lines derived from the dent, flint

and semi-dent parent lines, they were given numeric values of 0(dent), 1(flint), or 0.5(semi-dent). Least-squares means were calculated on a per line basis and used for all mapping.

Genotypic Data

Genotyping was done using Genotyping-By-Sequencing (GBS). The protocol for sequencing has been described [35]. Single nucleotide polymorphisms (SNPs) generated from GBS were used for association mapping. Synthetic intervals were created for the NAM population. Interval markers are derived from the common genotype within the interval. A 0.2cM interval size was imputed using a hidden Markov model described by Bradbury et al [37]. This results in 7,389 markers across the genome. These markers were used in linkage mapping, and joint linkage mapping.

Linkage Mapping

Composite interval mapping was used to map QTL in individual families in NAM. Composite interval mapping corrects for the linked marker bias added by traditional interval mapping [38] by using stepwise regression to fit a set of marker cofactors for each marker. It then uses these cofactors to calculate maximum likelihood to test putative intervals across the genome at regular intervals, not solely at marker locations. Support intervals were set to 1.5 LOD. This was done in the statistical computing environment R (http://www.r-project.org/), using the package R/QTL [39]. A threshold value for each family was calculated using 100 permutations at α = 0.05, to correct for the non-normal distribution of the data in each family. A 1 LOD interval around each significant peak was used for the confidence interval.

Joint Linkage Mapping

Joint linkage mapping was performed on the entire NAM population at once to give greater statistical power (due to larger sample size). Joint linkage mapping is more precise due to the increased number of recombination crossovers provided by the greatly increased sample size. This method controls for the population structure among the NAM founder lines. Joint linkage mapping uses a mixed model approach treating family and family by interval effects as fixed [27]. A stepwise regression model is fit over the data, and a refitting procedure is used to get rid of sites that do not improve the fit of the overall model [27]. Analysis was performed using TASSEL [40].

Association Mapping

Association mapping was used to map the correlation of sites to the phenotype. Unlike linkage mapping, association mapping exploits the recombination and mutation back several generations in history. Association mapping via genome wide association study (GWAS) was used in the maize association panel [32] to map significant SNPs. GWAS uses a Q + K mixed model, where genotypes and population structure are fixed effects (Q), line and kinship effects are random (K). This mixed model has been described [26]. The Genome Association and Prediction Integrated Tool (GAPIT) was used to carry out GWAS [41].

Candidate Genes

The metabolic pathways of starch synthesis and protein biosynthesis, as well as many other endosperm biochemical pathways, have been documented and continue to be studied. Many of the genes in these pathways are known, and many have been positionally cloned. Candidate genes' positions were obtained by searching the maize RefGen_v2 on www.maizegdb.org, and Maizecyc http://maizecyc.maizegdb.org/. Candidate genes were compiled and positions were cross referenced with support intervals for linkage mapping, joint linkage mapping, and GWAS.

Results

Phenotypic Distribution

The entire NAM population showed a wide distribution in kernel typse (Figure 1). B73, the common parent in this population has a dent kernel. Nine of the other founders are flint type, nine are semi-dent, and five are dent type kernels. The distribution seen in the entire NAM population is not normally distributed and is not even skewed toward the common parent (B73-dent). This distribution is heavily skewed to the flint type score. We can see that the association panel (Figure 2) mimics NAM in the distribution of kernel type, even though the two population were derived quite differently, (the NAM population is the result of a controlled crossing design, where the association panel lines were derived from numerous independent breeding efforts). Examination of the kernel type distribution by cross suggested non-additive modes of inheritance were common (Figure 3-5 Supplemental Figures S1-S19). For example, the dent by dent cross in Figure 6 does show that the dent type phenotype is recovered in the cross, but it is not common. There is a strong bias to recover more flint characteristics. The dent by flint cross in Figure 8 shows a similar distribution to that of the overall population in Figure 1.

Linkage Analysis

Composite interval mapping was used to assess each NAM family individually. In total, 25 QTL were found for an average of 1.1 loci per family (Table 1.) However five of the families produced no significant QTL. Out of the families that did produce QTL, there was an average of 1.5 loci per population. Results are shown in Table 1 and Figure 1. Many of the regions that are significant within a single family are found significant again in other families, or have overlapping confidence intervals if not exactly the same (see Supplementary Figures S20-S41 for individual family results). Candidate genes that are associated with each significant peak are listed in Table 1.

Results from the second form of linkage mapping, joint linkage mapping can also be seen in Figure 6, in the second row from the bottom, as well as in Table 2. Joint Linkage highlights 16 genomic regions of significance. Many of the joint linkage peaks are overlapping with the individual family linkage mapping studies previously mentioned (Figure 1). LOD scores in this experiment are greatly increased due to the large increase in sample used for the study (Table 2). Candidate genes that are associated with each significant result from the joint linkage mapping can be found in Table 2.

GWAS

To complement the NAM linkage mapping studies, GWAS was also conducted. The maize association panel [32] was used for this experiment. The genome wide Manhattan plot output from GAPIT [41] can be seen in Figure 7. This 282 line diversity panel has been used for many other GWAS studies [29, 30, 31, 33], but nothing in this study exceeds the significance threshold. While this analysis did not yield any statistically significant SNPs on its own, it does support many of the intervals already located in linkage and joint linkage mapping (Figure 6). Many of the significant points also cluster, so you can tell that SNPs in the same region are all responding to this test (Figure 6). There is no single outlier SNP in this plot.

Candidate Genes

Candidate genes that were considered in this experiment are listed in Table 3. There are 26 genes in storage protein biosynthesis and regulation related pathways. There are 48 candidate genes that belong to starch biosynthesis and regulation related pathways. In total, the candidate gene list used here is comprised of 74 genes. The positions of these genes were then cross-referenced with the positions of the significant markers in linkage and joint linkage (Table 1 and Table 2). Fourteen of the 74 genes line up with the regions of the genome described in Table 1 and Table 2. They are shown in Table 1, Table 2, and on Figure 6.

Discussion

The non-normal distribution of the data (Figure 1) is of interest. We would expect to see a relatively normal distribution based on the relatively balanced distribution of kernel types in the founding parents. As the common parent (B73) is dent type, one might expect the distribution to skew towards the dent type. But, in fact, we do not see this. Even a dent by dent cross (Figure 6) can result in some flint type kernels. The distribution of kernel type scores in the NAM population is heavily skewed towards the flint type kernels, even with a relatively balanced parental distribution (Figure 1). In the semi-dent by dent cross seen in Figure 4, again we get an extension of the scores into the flint end of the distribution with no purely flint parent in the cross. Also in Figure 5, with a dent by flint the distribution is not an average of the parental scores at all; instead, it is heavily skewed towards the flint parental type score. These distributions are similar throughout the NAM population (Supplemental Figures S1-S19). This suggests that this trait has a non-additive nature. It also suggests that this trait is controlled by many loci, which have threshold like expression.

There is some overlap of significant markers in the results from the joint linkage and individual family linkage mapping. Even though each family is different, they sometimes map to the same locations and they are still picking up on the same genomic regions as the combined joint linkage mapping. The LOD scores are much higher in the joint linkage experiment, which is to be expected due to a greater sample size, but the increase in LOD scores also suggests a continuity of the data and genetics across the entire population. This is confounded by the size of the confidence intervals. In Table 1 the median

confidence interval size is roughly 8 megabases, while in joint linkage (Table 2), the median confidence interval size is about 6 megabases, but the proportions again suggest that with joint linkage we are more confident in what we find.

Not only do we see that the linkage and joint linkage maps overlap for many of the significant regions but also for many of the candidate genes. The locations of many candidate genes are coinciding with significant markers in both of the linkage mapping studies. Thirteen of our 74 candidate genes in Table 3 were overlaid on Figure 6. These candidate genes are within 1 megabase of the confidence intervals of one of the peaks in linkage or joint linkage. These 13 genes are all in either starch or protein biosynthesis or regulation pathways in maize. In the joint linkage results, our confidence intervals sum up to about 8% of the genome. Our results highlight 13 of our 74 candidate genes which is 17% of the total candidate genes. Without controlling for recombination and gene density, we highlight double the number of genes we would by chance given the genome coverage of the joint linkage results. This shows that our results are enriched for our candidate genes. This gives us strong evidence that we are looking in the right place because both overlap, but also that we are hitting genes that we assume can affect the phenotype we are looking at.

The candidate genes that we find are interesting due to their activity in maize kernels. We find *Sucrose Synthase 1* (sus1) and *Sucrose Synthase 2* (*sus2*). These encode enzymes that are used to regulate the biosynthesis of sucrose. As sucrose is the precursor of starch, this could affect the assembly and structure of future starch granules. Similarly *Starch Branching Enzyme 1* (*sbe1*) and *Isoamylase type Starch Debranching Enzyme 3* (*iso3*) both regulate the branching patterns and granular shape of amylopectin. Also sus1 and *Shrunken 2* (*sh2*) are known mutants that cause an increase in sugars, a decrease in starch, and denting of the kernels (www.maizegdb.org).

On the protein side, we find a cluster of alpha-zein controlling genes. *Opaque 2* (*o2*), *Opaque2 Heterodimerizing Protein* (*oph2*), a *19KDa alpha-zein* (Z1D – referring to both *az19D1* and *az19D2* loci which are considered a cluster gene Z1D), and the *27KDa alpha-zein protein* (*zp27*) are all very closely interrelated genes that mostly work together to produce many of the alpha-zein storage proteins in the endosperm. *Opaque 2* is also a known endosperm mutant that on its own causes opaque and chalky kernels. Floury 1, 2, and 3 (*fl1*, *fl2*, *fl3*) are other genes that have known effects on kernel morphology. Mutants of these genes express with floury endosperm kernels that can be extremely fragile.

Even in the Maize Association Panel we see a similar outcome when testing for our kernel morphology phenotype. There appears to be a lot of noise, and no significant SNPs appear in the GWAS, but some of the genomic regions that are the most significant do support our already located regions of interest (Figure 7). We have a low sample number for this test (only n=282), which may not be enough to have the statistical power to find a SNP significant. Yet again we see similar regions being targeted by this study. Many of the mildly significant clusters of SNPs are localized to regions that were also highlighted by the candidate genes from the linkage and joint linkage mapping (Figure 4). This is more evidence that these genes may play a role in this phenotype.

We have mapped several QTL and associated candidate genes with this non-additively inherited complex trait. However, we cannot conclude that the candidate genes we have located are in fact the causal genes for this phenotype. More testing will be needed to confirm their role. We had certain candidate genes in mind prior to these mapping exercises, and a good portion of them lined up very well with our results. Our hits are still large genetics regions with many other genes in them, but our mapping results and GWAS results are good evidence that these genes are involved with the shaping of maize kernels. We have highlighted 13 of the 40,000 genes in maize as candidates for contribution to this phenotype. We now have a better idea of the genetics behind maize kernel morphology.

References

1. Hays MW (1890) Improving Corn- Cross Fertilization and Selection. Univ. of Minnesota Agric. Exp. Stm., St. Paul, MN. Bulletin 11.
2. Correns C (1901) Bastarde zwischen maisrassen, mit besonder Berucksichtung der Xenien. (In German.) Bibl. Botanica 53:1–161.
3. Tracy WF, Whitt SR, Buckler ES (2006) Recurrent mutation and genome evolution: example of *Sugary1* and the origin of sweet maize. *The Plant Genome* 46:S49-54.
4. Myers AM, James MG, Lin Q, Yi Q, Stinard PS, et al. (2011) Maize opaque5 encodes monogalactosyldiacylglycerol synthase and specifically affects galactolipids necessary for amyloplast and chloroplast function. The Plant Cell Online 23: 2331-2347.
5. Lin Q, Huang B, Zhang M, Zhang X, Rivenbark J, et al. (2012) Functional interactions between starch synthase III and isoamylase-type starch-debranching enzyme in maize endosperm. Plant physiology *158*: 679-692.
6. Collins G N (1909) A new type of Indian Maizefrom China. Bureau of Plant Industry Bulletin 161: 1-30.
7. Kempton JH (1919) Inheritance of waxy endosperm in maize. USDA Bull. 754.
8. Hayes HK, East EM (1915) Further experiments on inheritance in maize. Conn. Agric. Exp. Stn. Bull. 188: 1–31.
9. Nelson OE, Mertz ET, Bates LS (1965) Second mutant gene affecting the amino acid pattern of maize endosperm proteins. Science 150: 1469–1470.
10. Mertz ET, Nelson, OE, Bates LS (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. Science 145: 279–280.
11. Osborne TB (1891) Process of extracting zein. US Patent 456773.
12. Mertz ET, Bressani R (1957) Studies on maizeproteins I: A new method of extraction Cereal Chem., 34, pp. 63–69
13. Watson SA, Ramstad PE (1987) Corn: chemistry and technology. American Association of Cereal Chemists.
14. Delcour JA, Bruneel C, Derde LJ, Gomand SV, Pareyt B, et al. (2010) Fate of starch in food processing: from raw materials to final food products. Food Science and Technology 1 Doi:10.1146/annurev.food.102308.124211.

15. Keeling PL, and Myers AM (2010) Biochemistry and Genetics of Starch synthesis. Annu. Rev. Food Sci. Technol. 1: 271-303

16. Esen A (1987) A proposed nomenclature for the alcohol-soluble proteins (zeins) of maize (*Zea mays L.*). J Cereal Sci 5: 117–128.

17. Matsushima N, Danno G, Takezawa H, Izumi Y (1997) Three-dimensional structure of maize α-zein proteins studied by small-angle X-ray scattering, Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1339: 14-22.

18. Wu Y, Holding DR, Messing J (2010) γ-Zeins are essential for endosperm modification in quality protein maize. Proceedings of the National Academy of Sciences 107: 12810-12815.

19. Argos O, Pedersen K, Marks MD, Larkins BA (1982) A Structural Model for Maize Zein Proteins. J. Biol. Chem. 257: 9984–9990.

20. Holding DR, Otegui MS, Li B, Meeley RB, Dam T, et al. (2007) The maize floury1 gene encodes a novel endoplasmic reticulum protein involved in zein protein body formation. The Plant Cell Online 19: 2569-2582.

21. Wallace HA, Brown WL (1956) Maizeand its early fathers. Maizeand its early fathers.

22. Brown WL, Zuber MS, Darrah LL, Glover DV (1985) The National MaizeHandbook: Origin, Adaptation, and Types of Corn. Iowa State University Cooperative Extension Service.

23. Sprague GF (1955) Maizeand MaizeImprovement. New York Academic Press.

24. Gustafson T, de Leon N (2010) Genetic analysis of maize (*Zea mays* L.) endosperm vitreousness and related hardness traits in the Intermated B73 x Mo17 recombinant inbred line population. *Crop Science* 50: 2318-2327.

25. Van der schaaf P, Wilbur DA, Painter RH (1969) Resistance of Maizeto Laboratory Infestation of the Larger Rice Weevil, *Sitophiltts zeamai*. Journal of Economic Entomology. 62: 352-355.

26. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**: 203-208.

27. Buckler ES, Holland JB, Bradbury PJ, Acharya C, Brown P et al. (2009) The genetic architecture of maize flowering time. *Science* **325**: 714-718.

28. McMullen MD, Kresovich S, Villeda HS, Bradbury PJ, Li H et al. (2009) Genetic properties of the maize nested association mapping population. *Science* **325**: 737-740.

29. Tian F, Bradbury PJ, Brown PJ, Sun Q, Flint-Garcia S et al. (2011) Genome-wide association study of maize identifies genes affecting leaf architecture. *Nature Genetics* **43**: 159-162.

30. Cook JP, McMullen MD, Holland JB, Tian F, Bradbury PJ, et al. (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology* **158**: 824-834.

31. Larsson SJ, Lipka AE, Buckler ES (2013) Lessons from Dwarf8 on strengths and weaknesses in structured association mapping. *PLoS Genetics* **9**: e1003246.

32. Flint-Garcia SA, Thuillet A-C, Yu JM, Pressoir G, Romero SM et al. (2005) Maize association population: a high resolution platform for quantitative trait locus dissection. *Plant Journal* **44**: 1054-1064.

33. Kump KL, Bradbury PJ, Buckler ES, Belcher AR, Oropeza-Rosas M et al. (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nature Genetics* **43**: 163–168.

34. Youens-Clark K, Buckler ES, Casstevens T, Chen C, DeClerck G et al. (2010) Gramene database in 2010: updates and extensions. *Nucleic Acids Research* **39**: D1085–D1094.

35. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6: e19379.

36. Chia JM, Song C, Bradbury PJ, Costich D, de Leon N (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genetics* **44**: 803-807.

37. Bradbury et al. (in prep).

38. Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. Genetics 175: 361-374.

39. Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889-890.

40. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y et al. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633-2635.

41. Lipka AE, Tian F, Wang Q, Peiffer J, Le M et al. (2012) GAPIT: Genome Association and Prediction Integrated Tool. *Bioinformatics* **28**: 2397-2399.

Table 1. Summary of results from Independent family linkage mapping.

Population	LOD Scores	Threshold	Chromosome	Position (bp)	Lower Confidence interval (bp)	Upper Confidence Interval (bp)	Interval size (bp)	Candidate Gene
1	5.4	5.35	1	44044791	40476001	45513415	5037414	sus2
1	7.15	5.35	5	49897470	28723055	65275605	36552550	sbe1
2	7.32	5.15	2	23667776	20081914	28462548	8380634	
3	8.41	5.18	2	20510509	20040930	24049806	4008876	
4	5.96	5.27	2	196125696	195188506	197482927	2294421	
4	6.18	5.27	8	107232539	100566541	150464488	49897947	fl3
5	5.6	5.26	1	72629352	66429994	84195403	17765409	sus2
5	6.09	5.26	8	146365593	145371163	151527905	6156742	
6	6.56	5.11	8	114028492	102108063	122444198	20336135	fl3
7	6.02	5.11	5	212682814	212192544	212947219	754675	GRMZM2G419257_P01
8	5.45	5.21	3	24659629	21682687	107873101	86190414	
8	5.63	5.21	9	114291284	110404915	116482576	6077661	sus1
9	6.18	4.94	3	224008575	223397675	225454289	2056614	sh2
9	8.27	4.94	5	56737096	41138800	76634666	35495866	sbe1
12	7.24	4.64	8	171753107	171555914	172550484	994570	
13		6.69						
14	5.7	5.16	3	172587424	171444374	173596102	2151728	
15	8.35	5.53	1	62861084	56268023	64969651	8701628	sus2
16		5.44						
18	9.87	5.28	2	50634516	37448690	53522615	16073925	fl1
18	6.3	5.28	4	67852994	36356063	139144750	102788687	
19	5.97	5.25	1	301242248	296767429	301242248	4474819	
20		5.3						
21	8.5	5.44	4	39862418	31324943	43214166	11889223	
21	5.77	5.44	8	24721886	20755375	86708840	65953465	
22	4.72	4.72	3	219198452	218120392	221614870	3494478	sh2
22	4.85	4.72	4	223907294	216225151	227422943	11197792	
23		5.58						
25		5.2						
26	6.37	5.72	7	11616794	10463037	15238681	4775644	o2
Average	6.53	5.30				Median	8380634	

Table 2. Summary results from joint linkage mapping in the entire NAM population

Chromosome	LOD Scores	Position (bp)	Lower Confidence interval (bp)	Upper Confidence Interval (bp)	Interval size (bp)	Candidate Gene
1	6.83	149357300	101766503	164816528	63050025	sus2
1	7.73	207791849	200621829	209337816	8715987	Z1D
1	7.28	51327581	49944220	52056768	2112548	
2	16.04	49846838	42336272	66952853	24616581	fl1
2	7.6	209581871	209084703	210456683	1371980	
3	9.49	10867998	10261184	12986293	2725109	
3	7.48	219098578	216093032	222777164	6684132	sh2
3	8.46	170726091	167422420	172015899	4593479	
4	9.81	18478819	18150821	23568385	5417564	fl2
4	8.11	159358250	147403300	167066431	19663131	
5	9.23	80166258	73833475	83712444	9878969	
5	13.1	213520096	213167556	214489581	1322025	GRMZM2G419257_P01
5	9.47	10104367	7527027	13300783	5773756	
7	8.5	126349286	121950424	130705336	8754912	zp27, iso3
8	29.7	109794778	109154219	111716456	2562237	fl3
9	13.2	105831872	88651842	113606794	24954952	sus1
Average	10.75			Median	6228944	

Table 3. Candidate genes that were assessed in this study. Locus name, chromosome and related pathway are given for 75 candidate genes.

locus	chr	pathway		Locus	chr	pathway		locus	chr	pathway
az19D1	1	Protein		dzs10	9	Protein		sbe1b	10	Starch
az19D2	1	Protein		GRMZM2G070172_P01	1	Starch		GRMZM2G126988_P01	5	Starch
fl1	2	Protein		GRMZM2G163437	1	Starch		GRMZM2G419257_P01	5	Starch
hfi1	2	Protein		GRMZM2G311182_P01	1	Starch		sbe1	5	Starch
mc1	2	Protein		GRMZM2G391936	1	Starch		sh4	5	Starch
pbf1	2	Protein		sus2	1	Starch		GRMZM2G027955	6	Starch
zpu1	2	Protein		amya3	2	Starch		GRMZM2G169073_P01	6	Starch
az22z1	4	Protein		gpm120	2	Starch		iso2	6	Starch
az22z3	4	Protein		GRMZM2G076508_P01	2	Starch		sbe1a	6	Starch
dzr1	4	Protein		GRMZM2G106213_P01	2	Starch		su2	6	Starch
fl2	4	Protein		GRMZM5G897776_P01	2	Starch		GRMZM2G008263_P01	7	Starch
o1	4	Protein		GRMZM5G863596_P01	3	Starch		GRMZM2G074781_P01	7	Starch
zp1	4	Protein		sh2	3	Starch		GRMZM2G081502_P01	7	Starch
zpl1a	4	Protein		bt2	4	Starch		GRMZM2G144002_P01	7	Starch
zpl1b	4	Protein		GRMZM2G045171_P01	4	Starch		iso3	7	Starch
zpl1d	4	Protein		GRMZM2G068506	4	Starch		GBSS1	7	Starch
zpl1e	4	Protein		GRMZM2G130043_P01	4	Starch		GRMZM2G044744_P01	8	Starch
zpl1f	4	Protein		su1	4	Starch		GRMZM2G061795_P01	8	Starch
ohp2	5	Protein		ae1	5	Starch		sbe3	8	Starch
glb3	6	Protein		bt1	5	Starch		sh1	9	Starch
zp15	6	Protein		GRMZM2G046117_P01	5	Starch		ss1	9	Starch
gz50	7	Protein		GRMZM2G060659_P02	5	Starch		sus1	9	Starch
o2	7	Protein		GRMZM2G103055_P01	5	Starch		wx1	9	Starch
zp27	7	Protein		GRMZM2G105791_P01	5	Starch		du1	10	Starch
fl3	8	Protein		GRMZM2G121612_P01	10	Starch				

Figure 1. Distribution of kernel type scores in the entire NAM population. Kernel type score is the x-axis and the frequency of each type is on the y-axis.

Frequency of Kernel Type in Nam
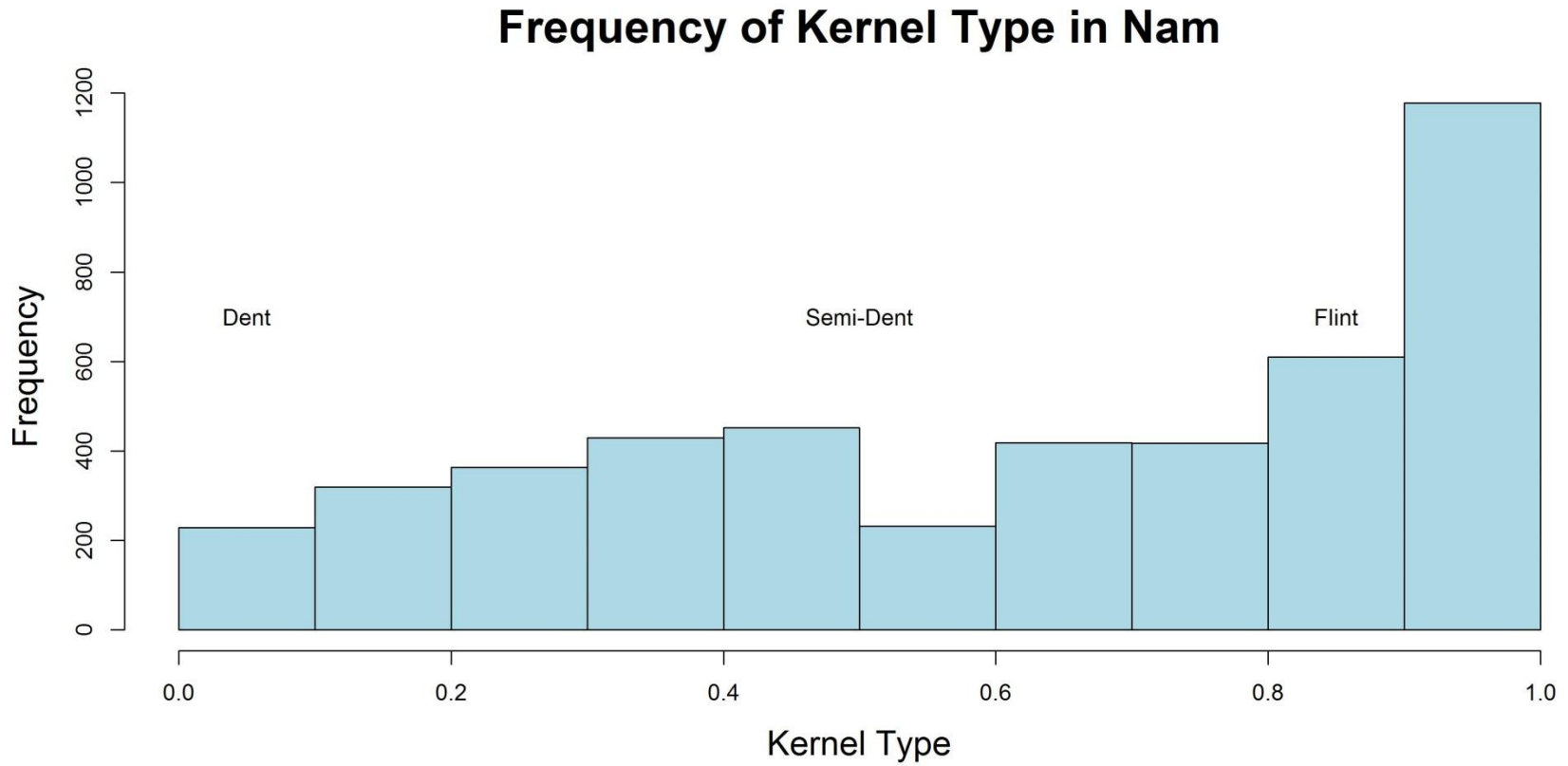
Figure 2. Frequency of kernel type score in the maize association panel. Kernel type score is the x-axis and the frequency of each type is on the y-axis.

Frequency of Kernel Type in Association Panel
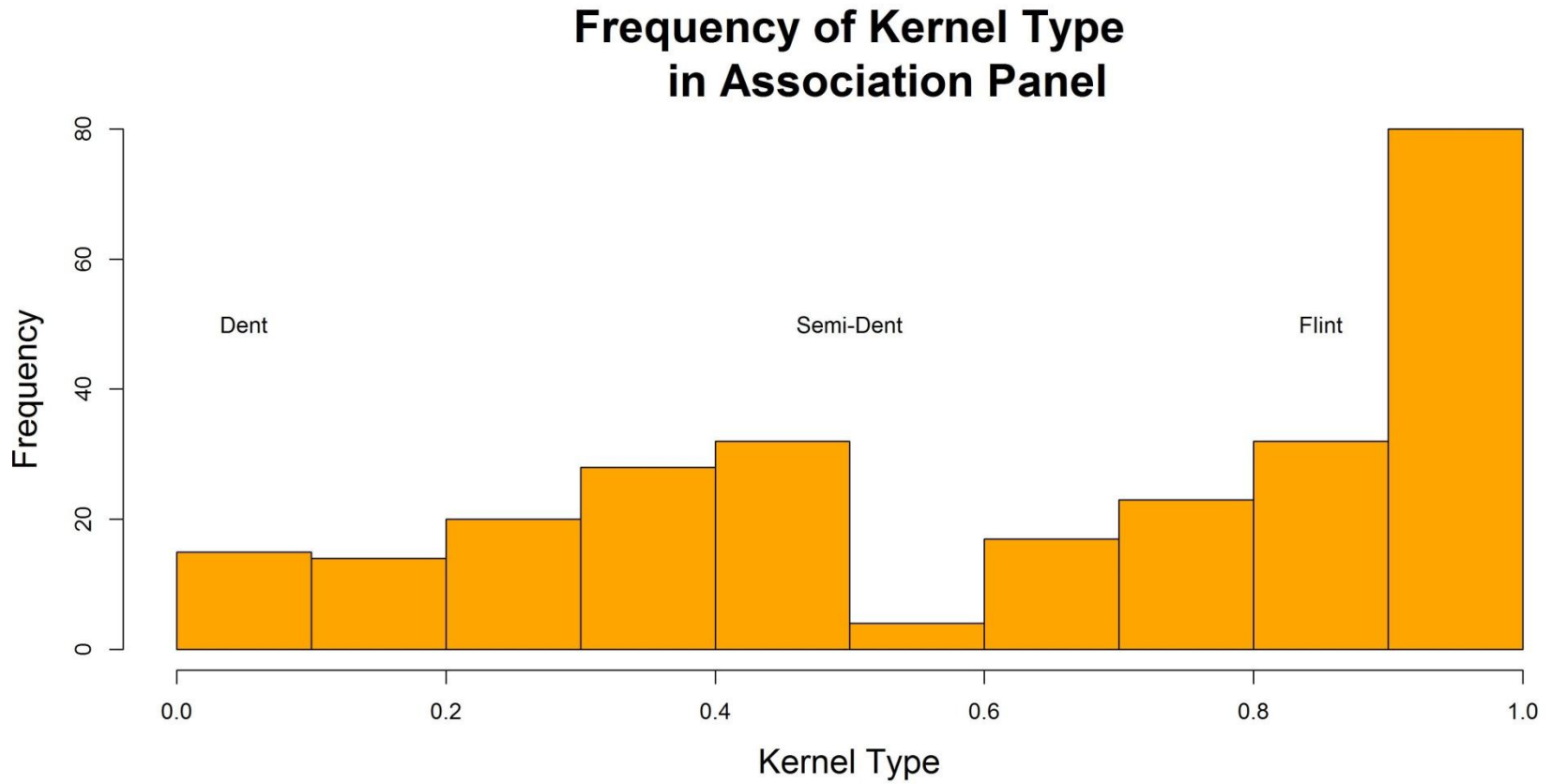
Figure 3. Population 1 of NAM, B73 x B97: A dent x dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
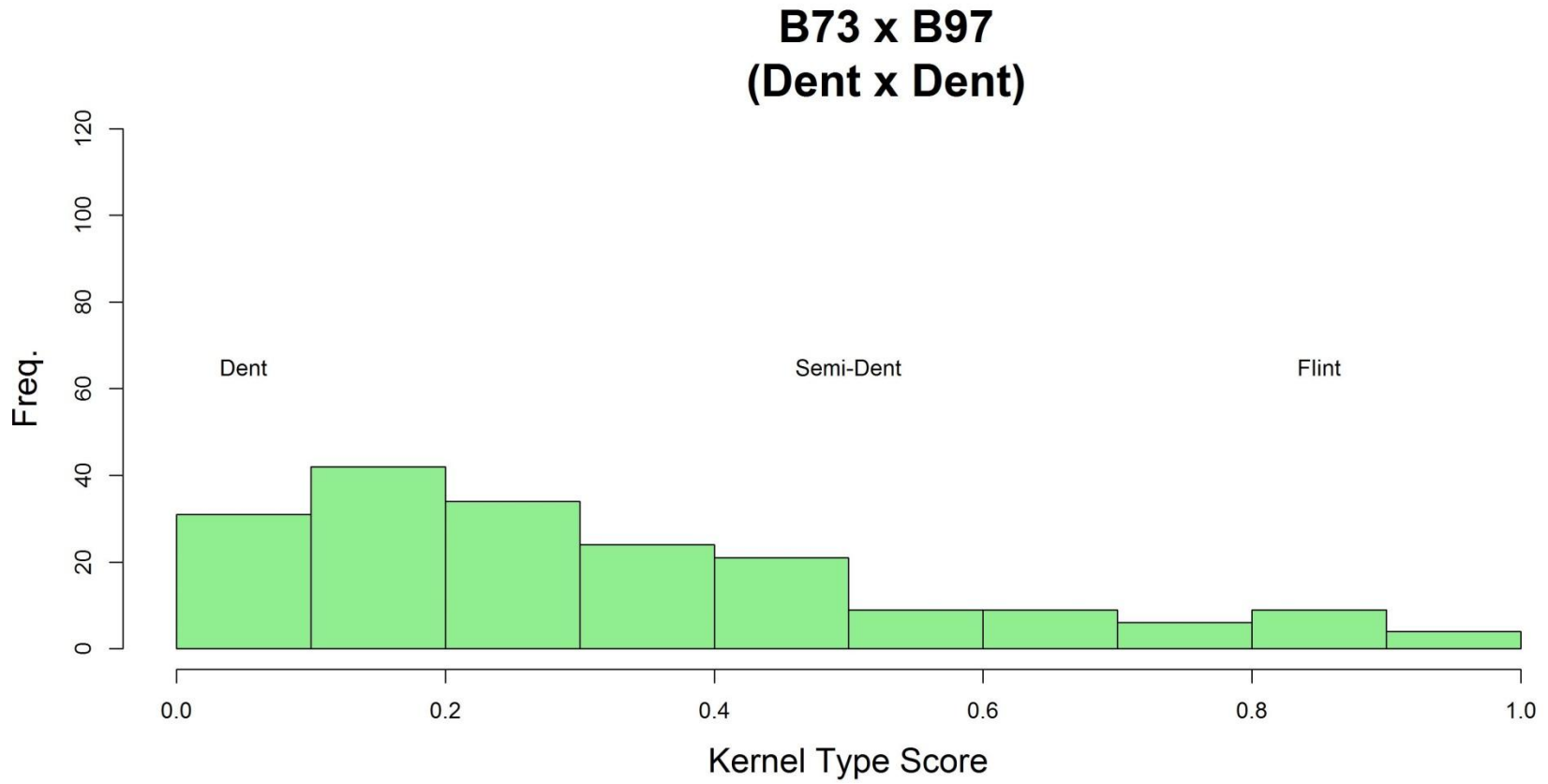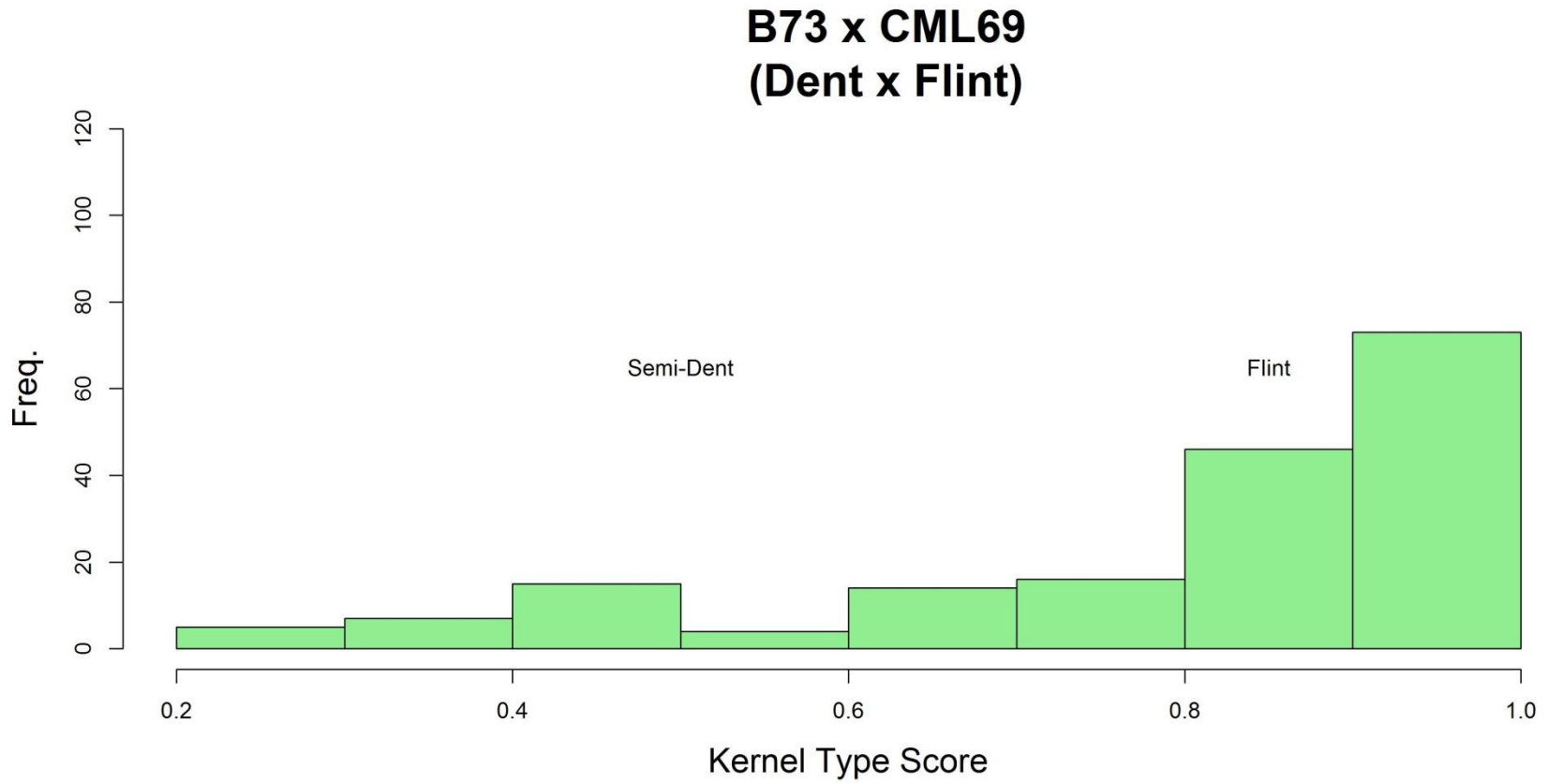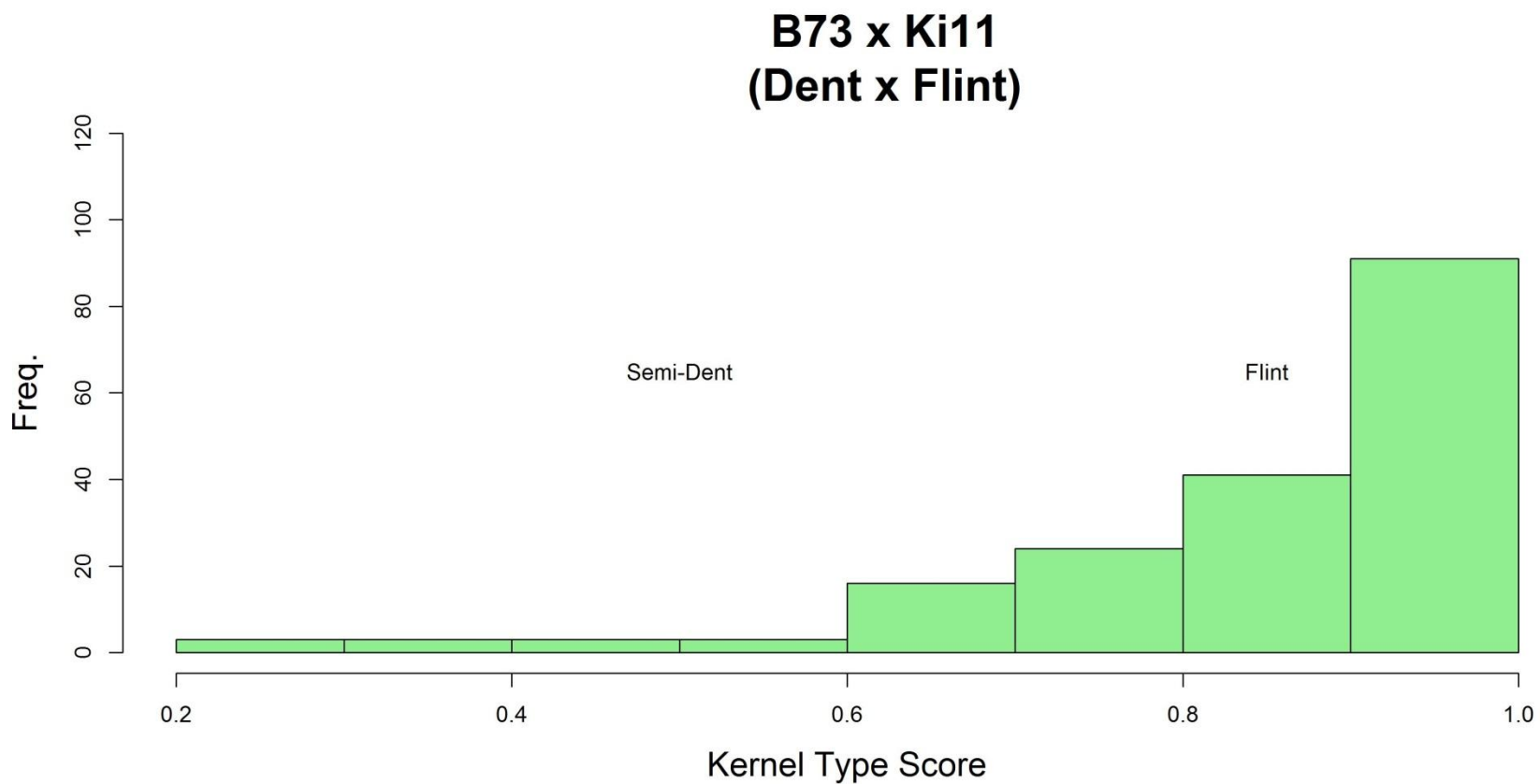
Figure 4. Population 15 of NAM, B73 x M162w: A dent x semi-dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
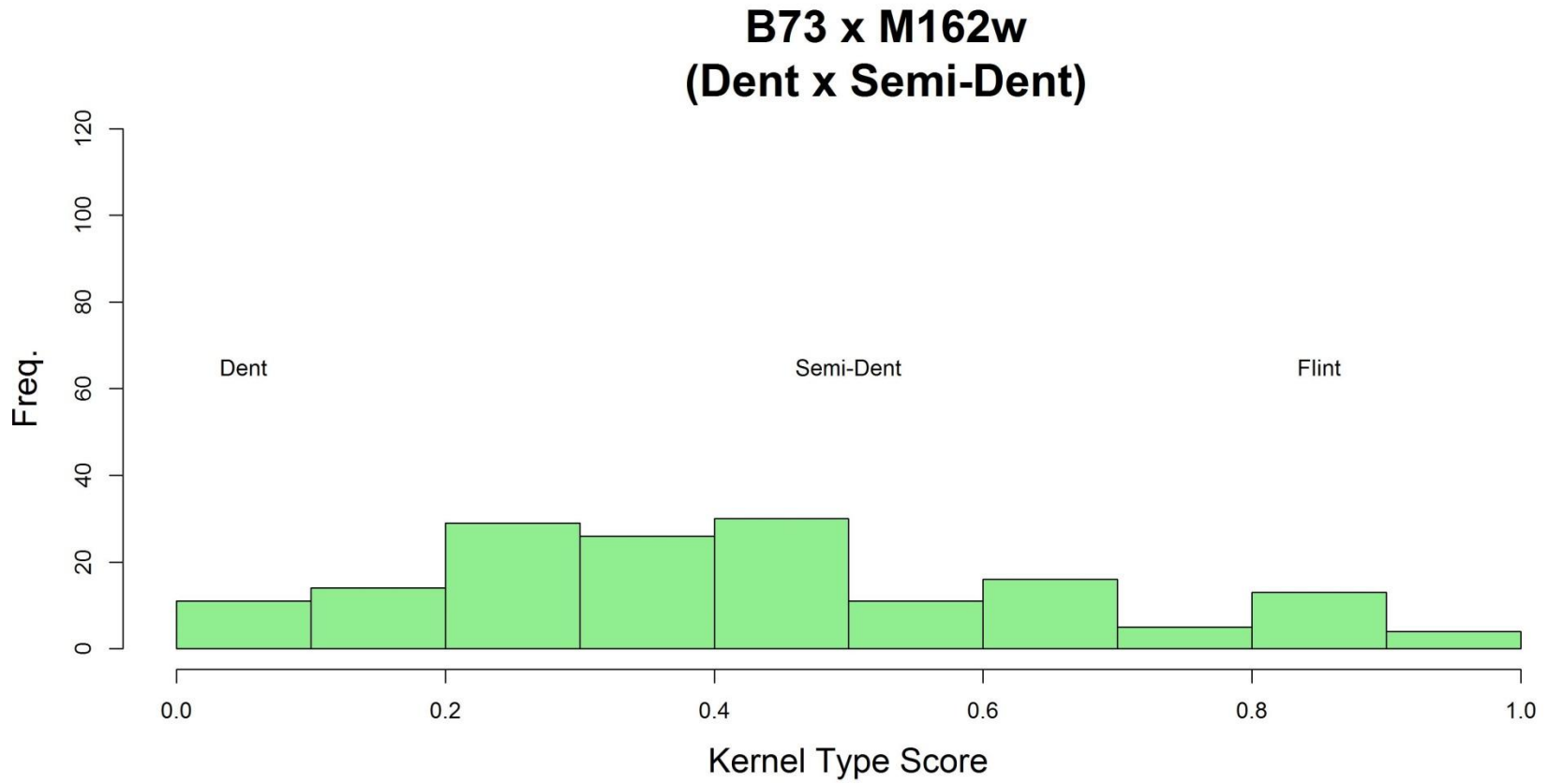
Figure 5. Population 7 of NAM, B73 x CML333: A dent x flint cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
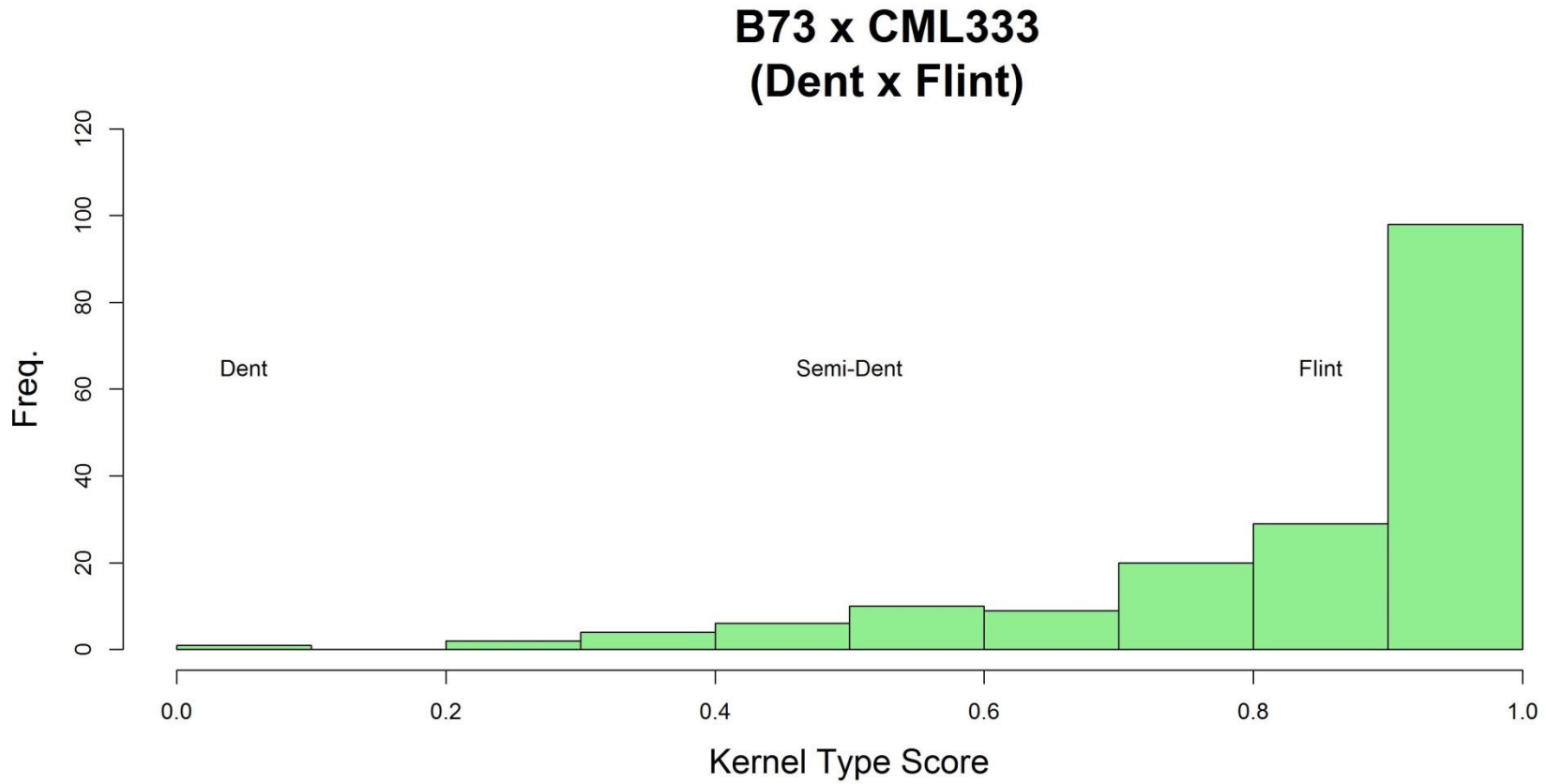
Figure 6. Summary of the mapping studies.Positon across the genome is on the x-axis, by chromosome. The y-axis contains each family of the NAM population as well as joint linkage (second to bottow row) and GWAS (bottow row).Grey lines are chromosome dividers. Each point is a peak from an indidual study. The blue lines are confiodence intervals calculated at 1 LOD. Every significant peak found in individual linkage or joint linkage mapping is shown here. Any SNP with a –log10 p-value was added here.Red lines are positions of the candidate genes from Table 1 and Table 2.

Summary of Mapping Studies
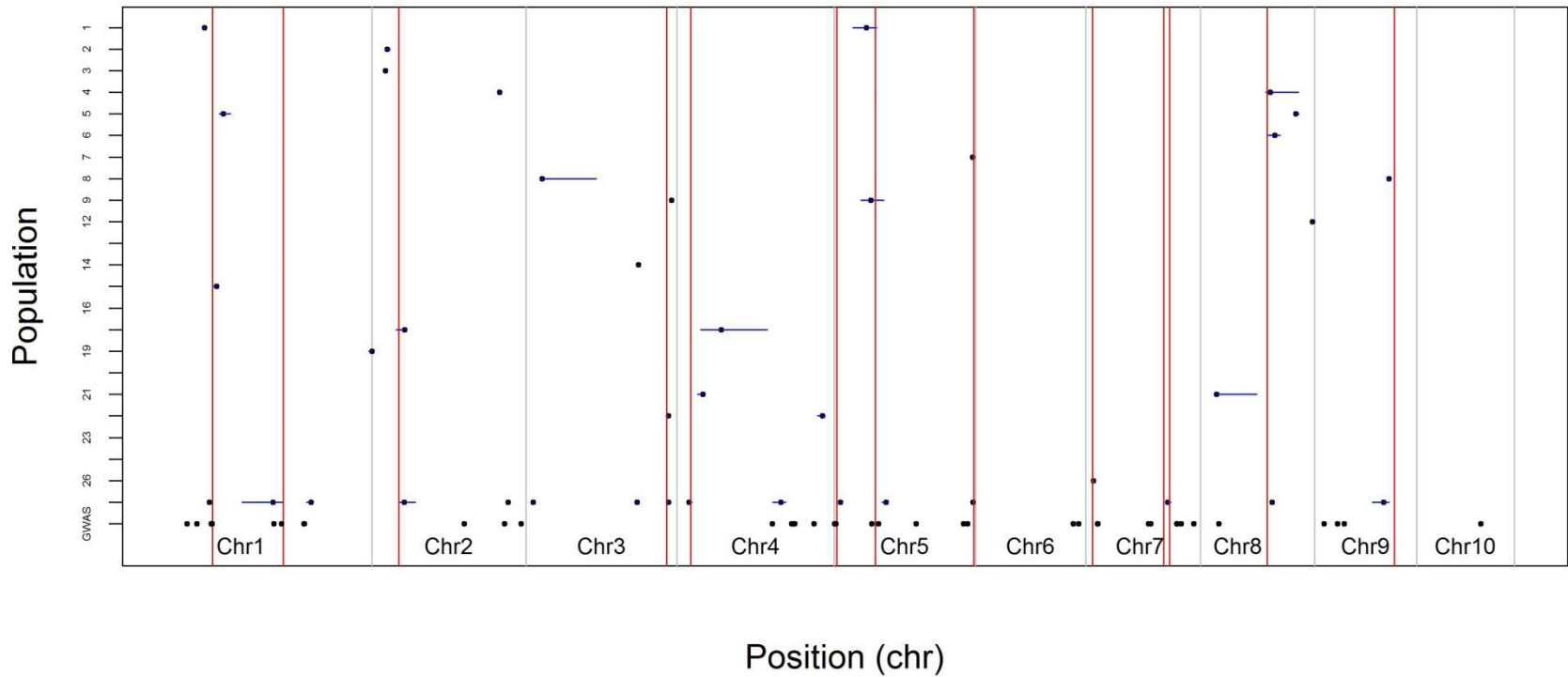
Figure 7. Results from GWAS of the maize assocaition panel. Genomic position by chromosome is on the x-axis, and –log10 p-values are given on the y-axis. No SNPs break the threshold. Peaks start to line up with linakage and joint linakage peaks (Figure 1).

Genome Wide Association Study results from the maize association panel
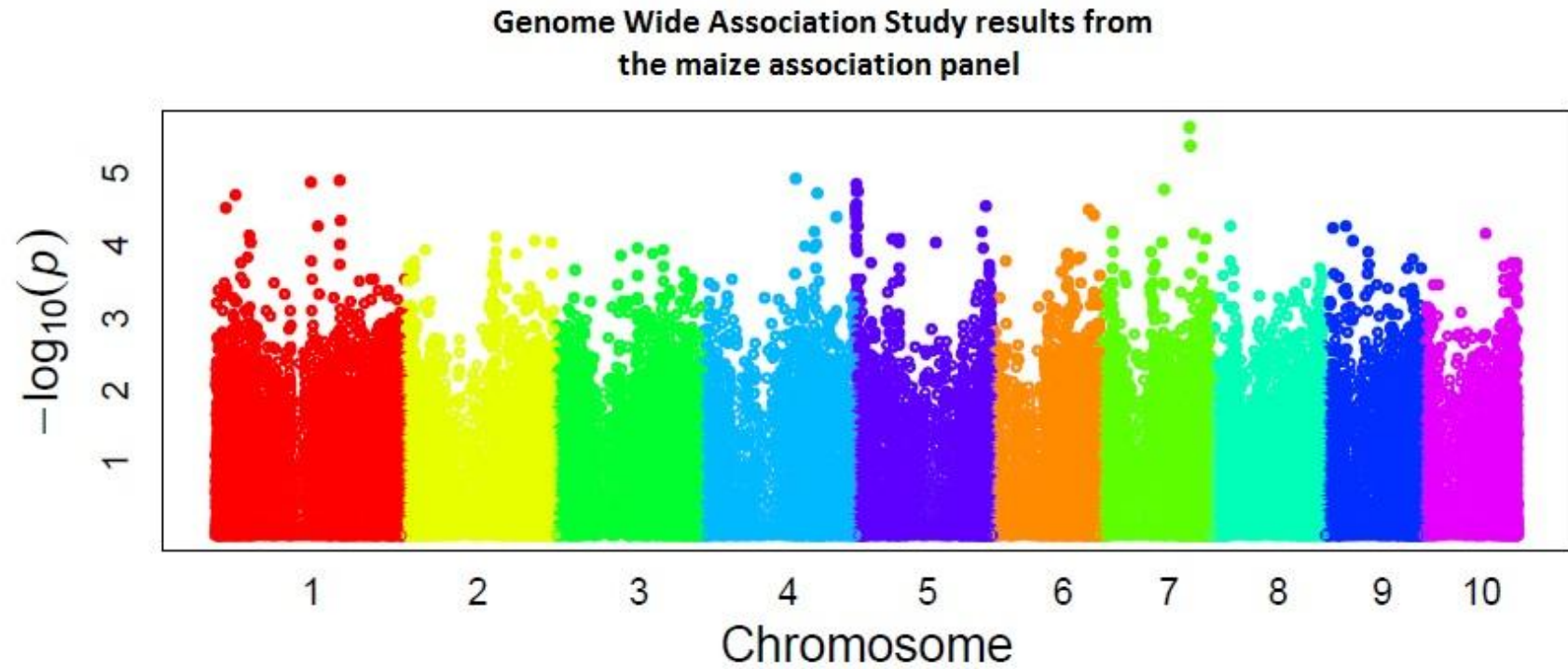
Supplemental Figures.

Figure S1. Population 2 of NAM, B73 x CML103: A dent x semi-dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
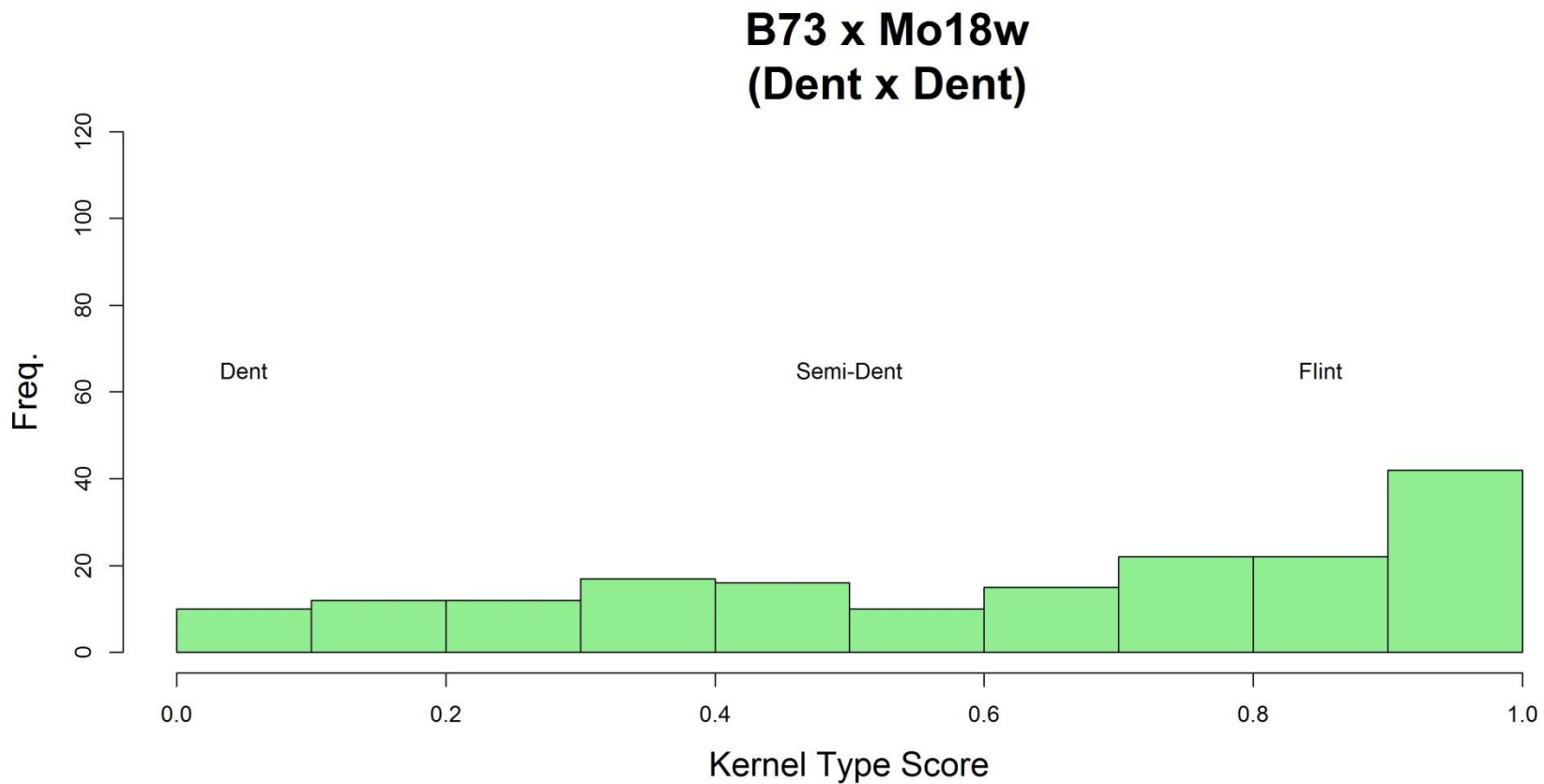

**B73 x CML103
(Dent x Semi-Dent)**
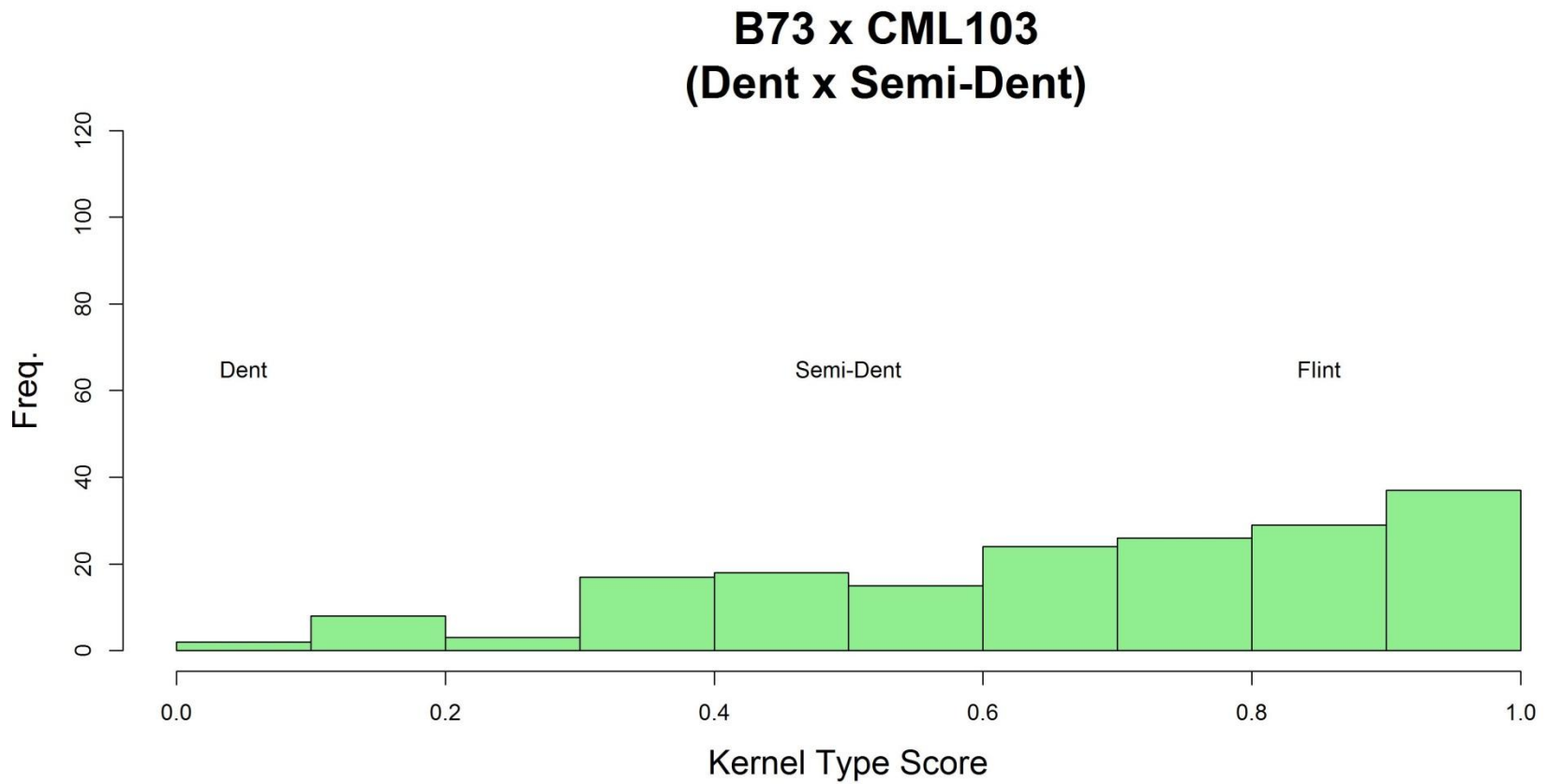
Figure S2. Population 3 of NAM, B73 x CML228: A dent x flint cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
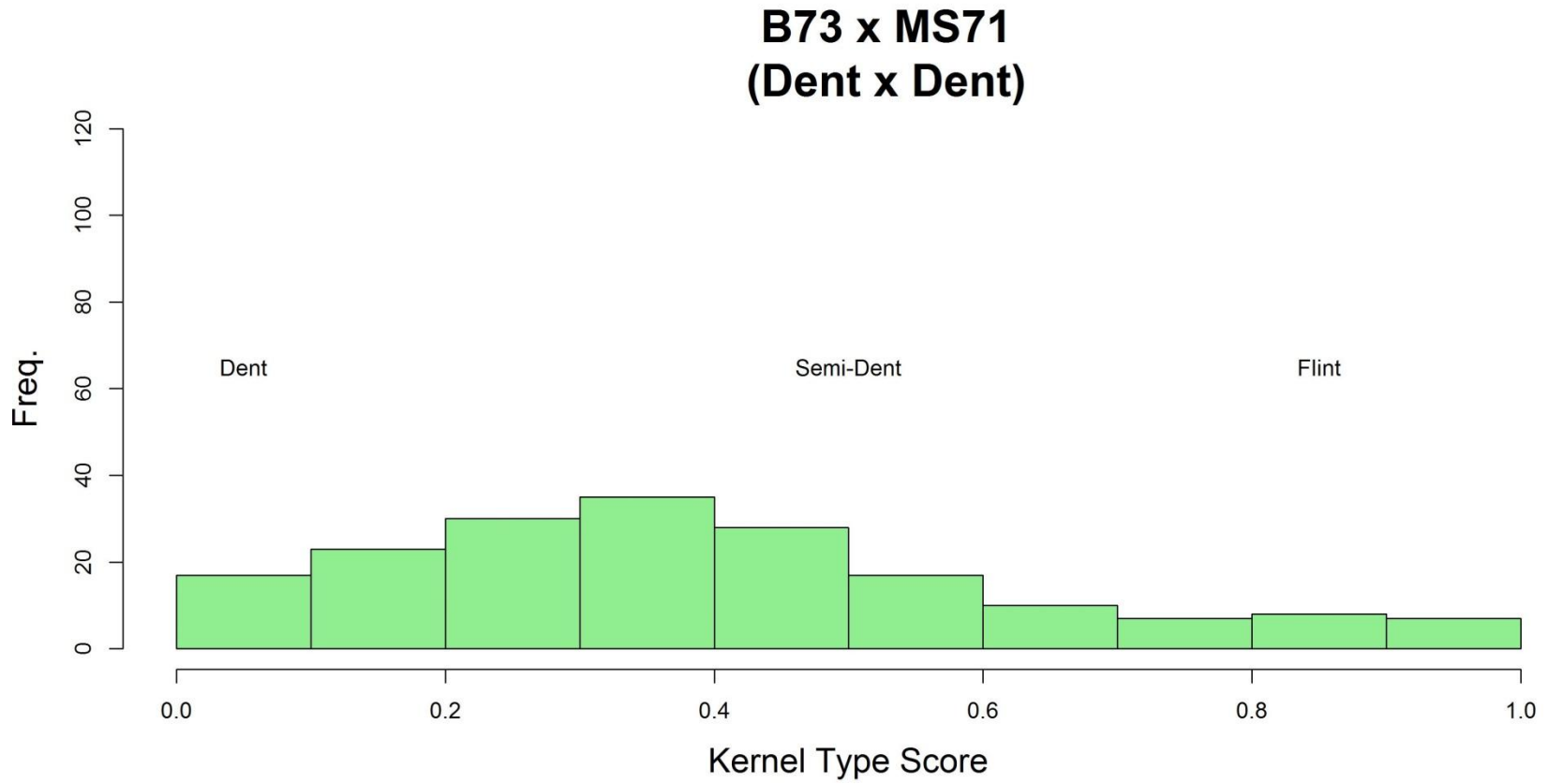

B73 x CML228
(Dent x Flint)
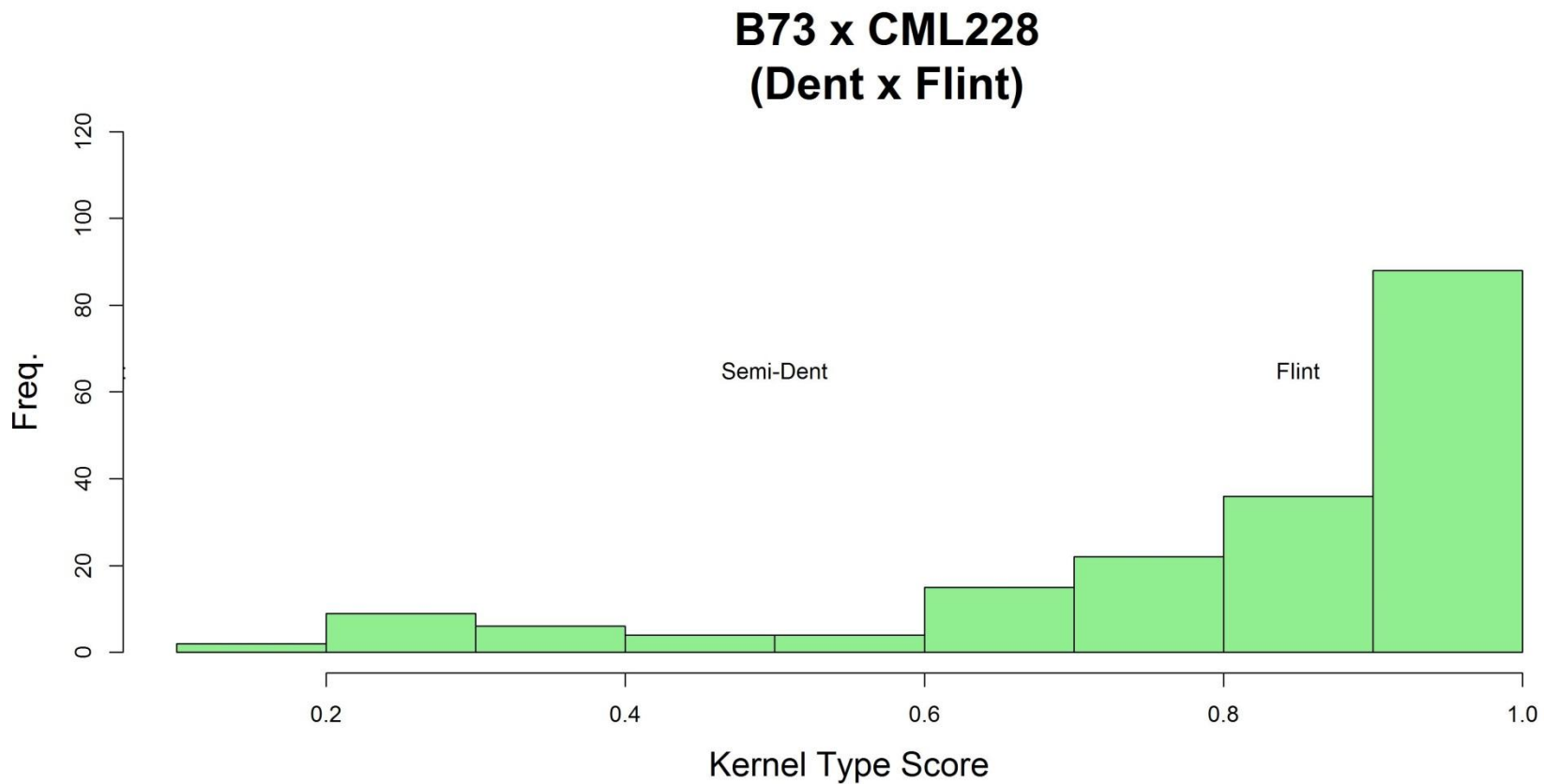
Figure S3. Population 4 of NAM, B73 x CML247: A dent x dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
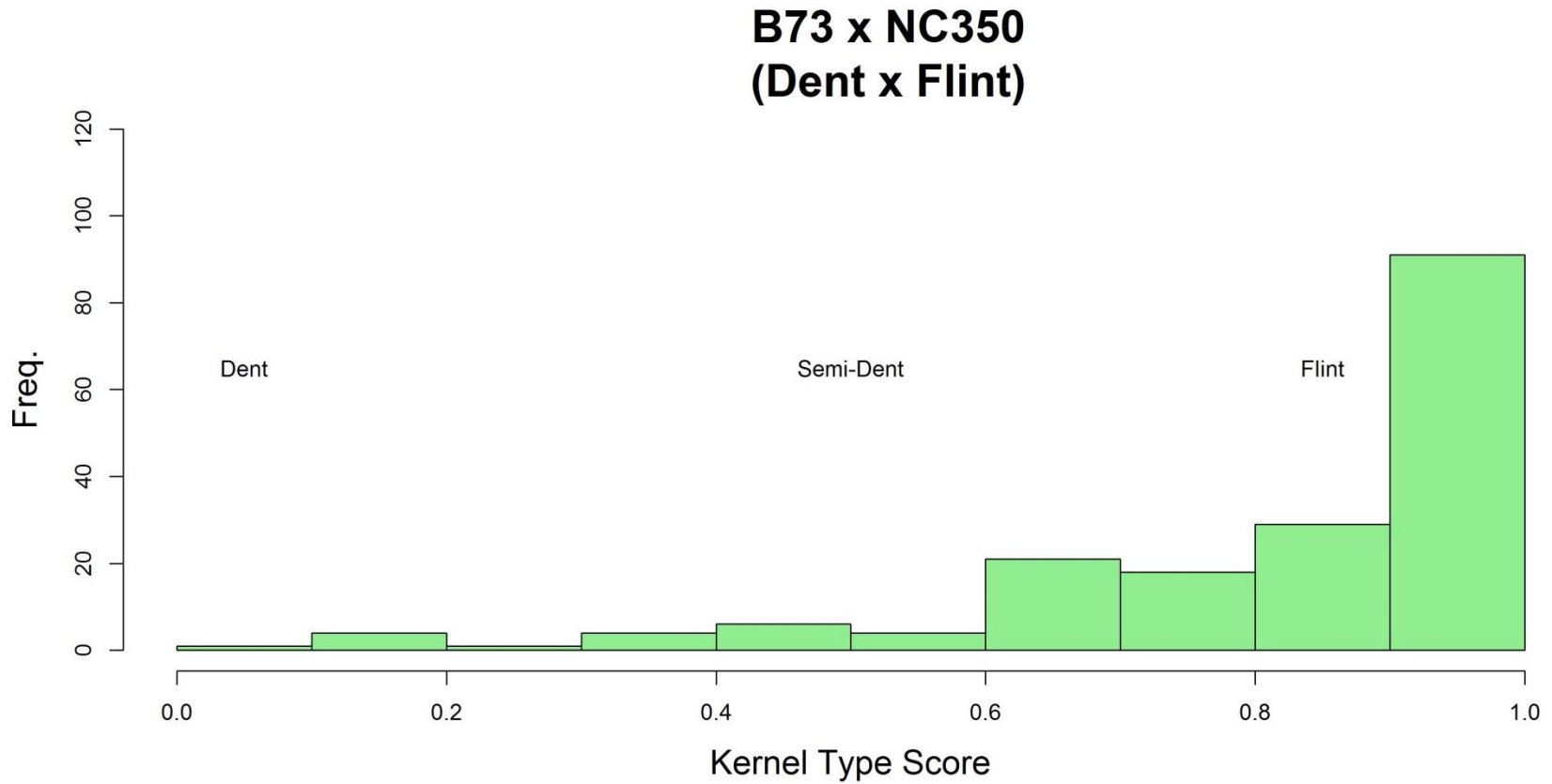

**B73 x CML247
(Dent x Dent)**
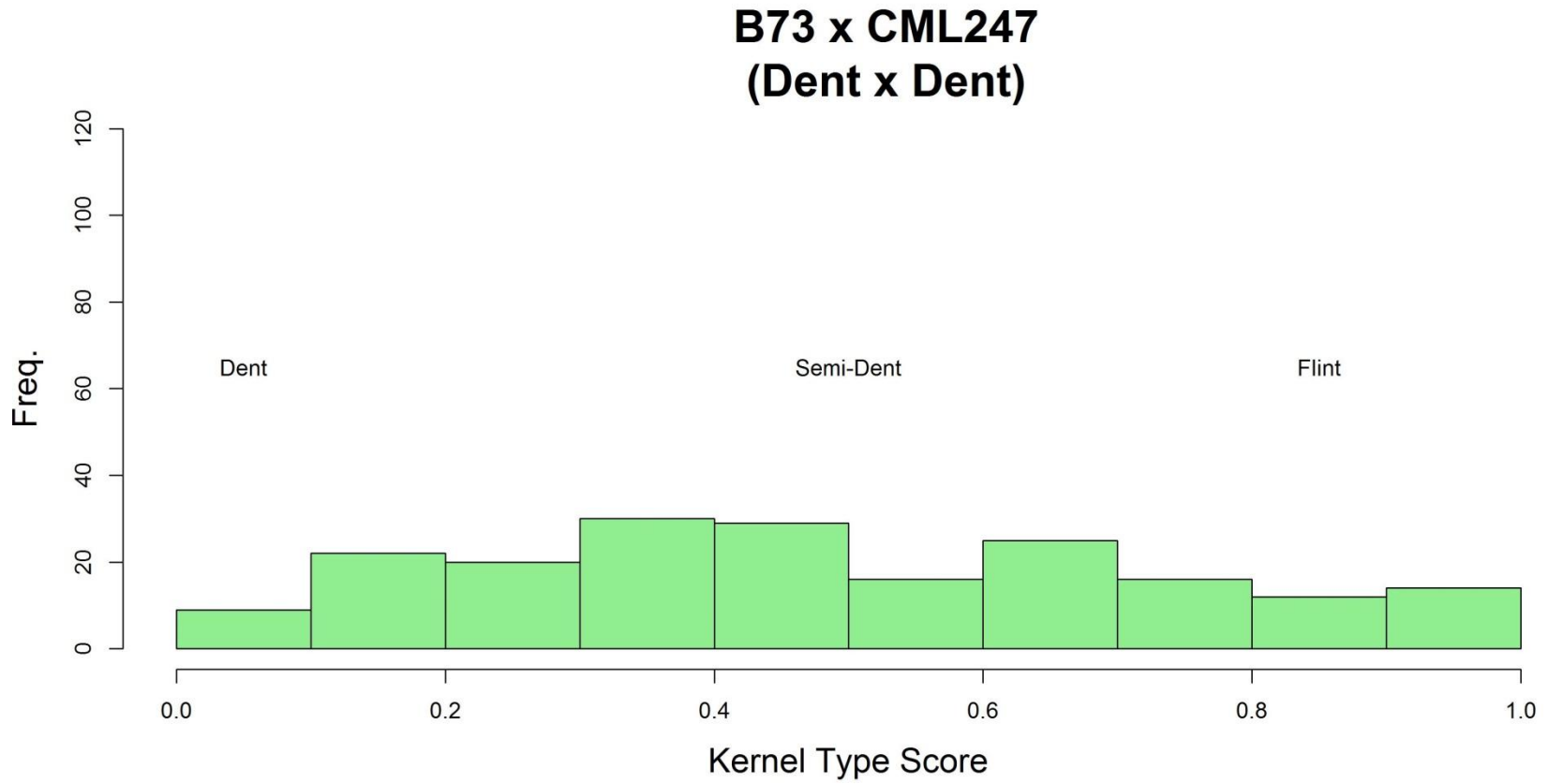
Figure S4. Population 5 of NAM, B73 x CML277: A dent x flint cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

B73 x CML277
(Dent x Flint)
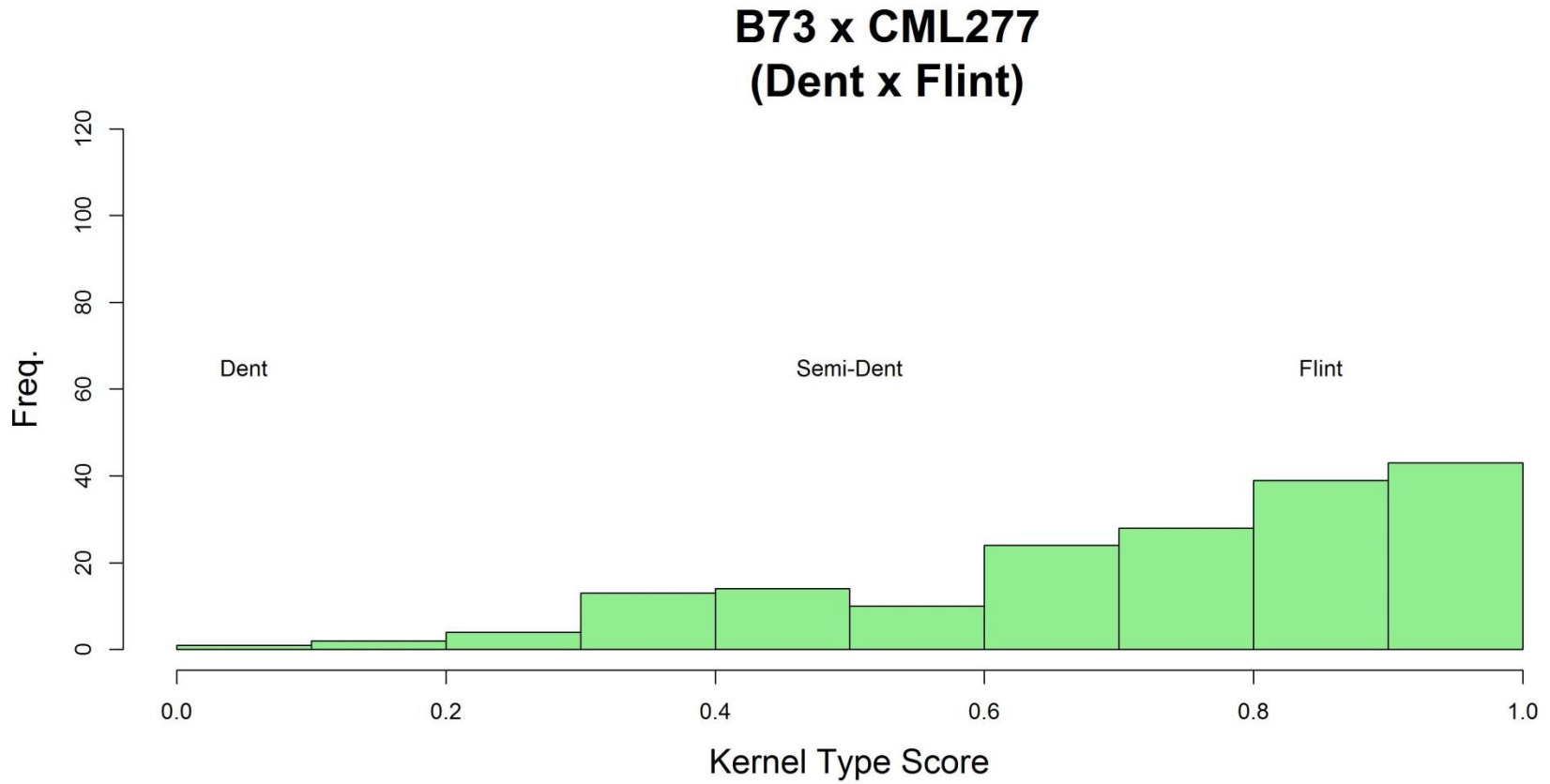
Figure S5. Population 6 of NAM, B73 x CML322: A dent x semi-dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
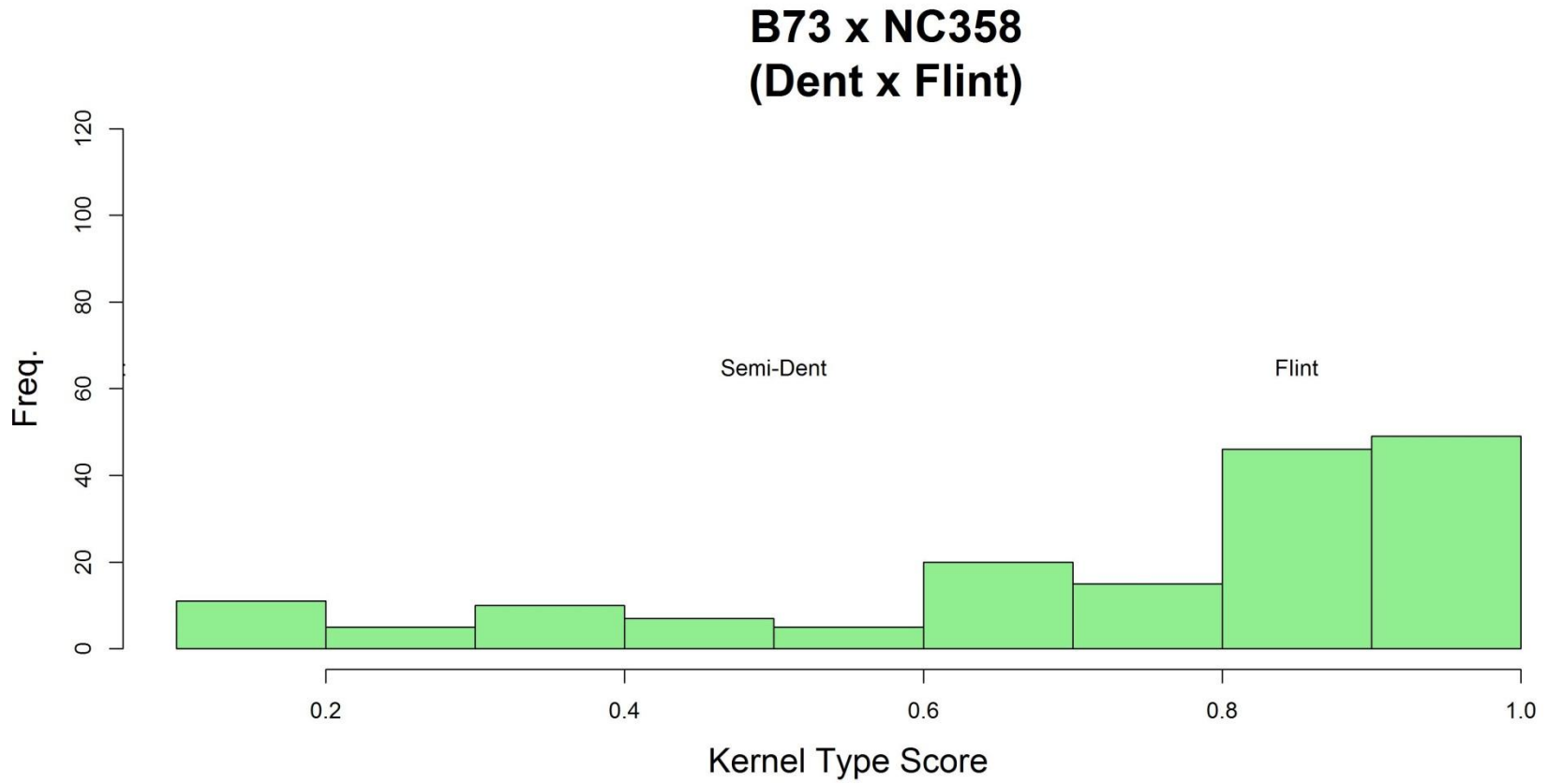

**B73 x CML322
(Dent x Semi-Dent)**
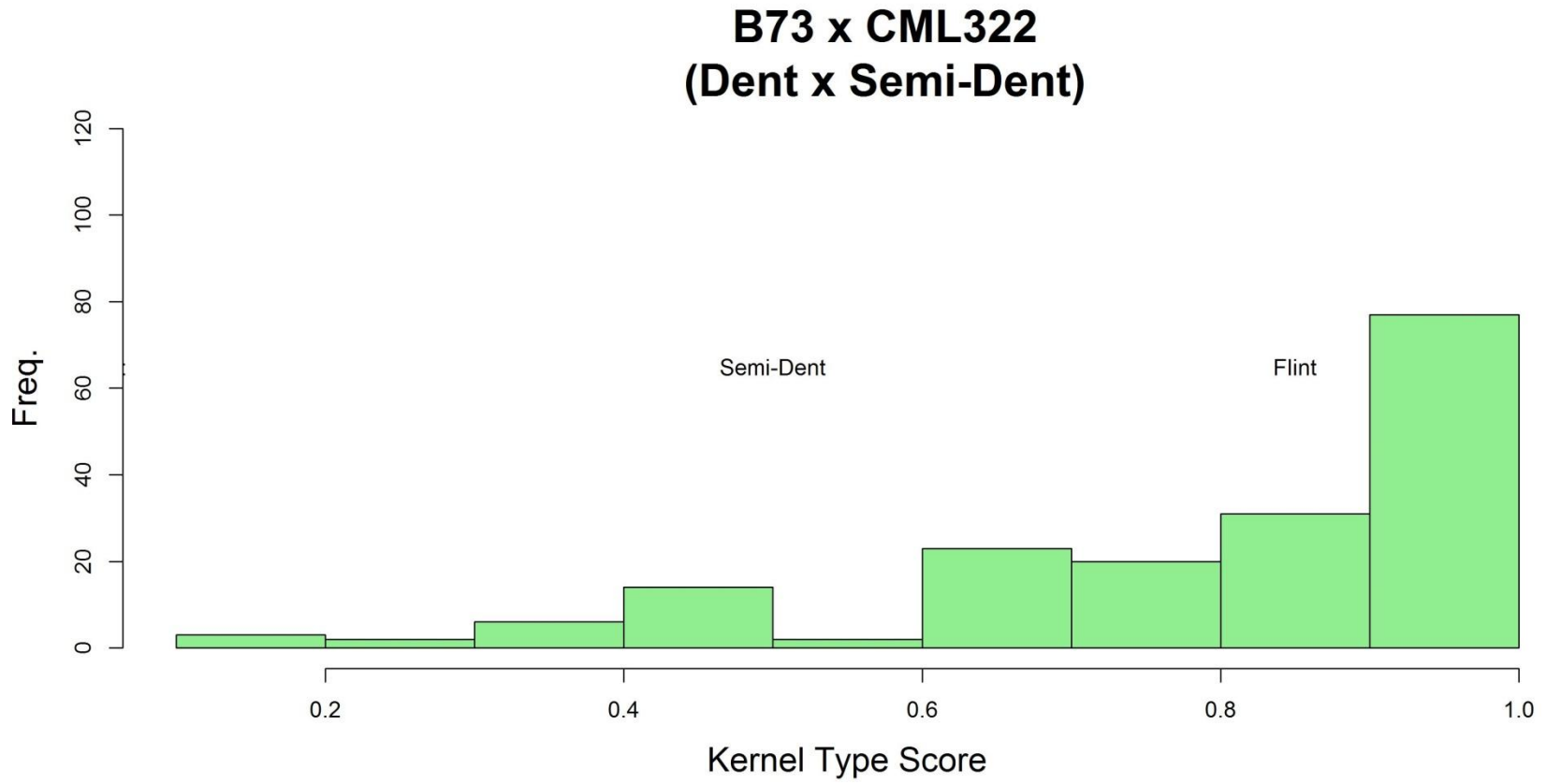
Figure S6. Population 8 of NAM, B73 x CML52: A dent x flint cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
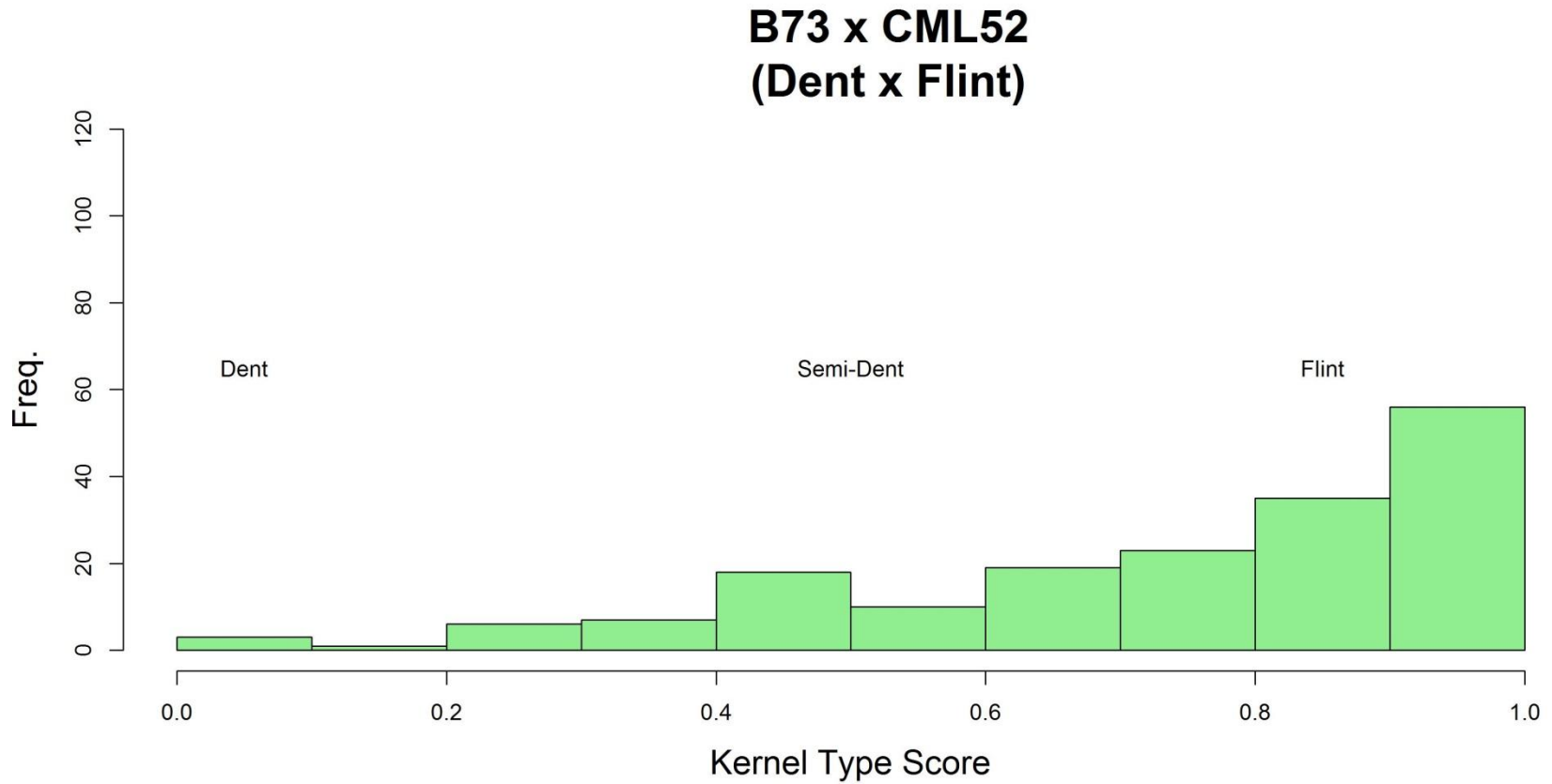
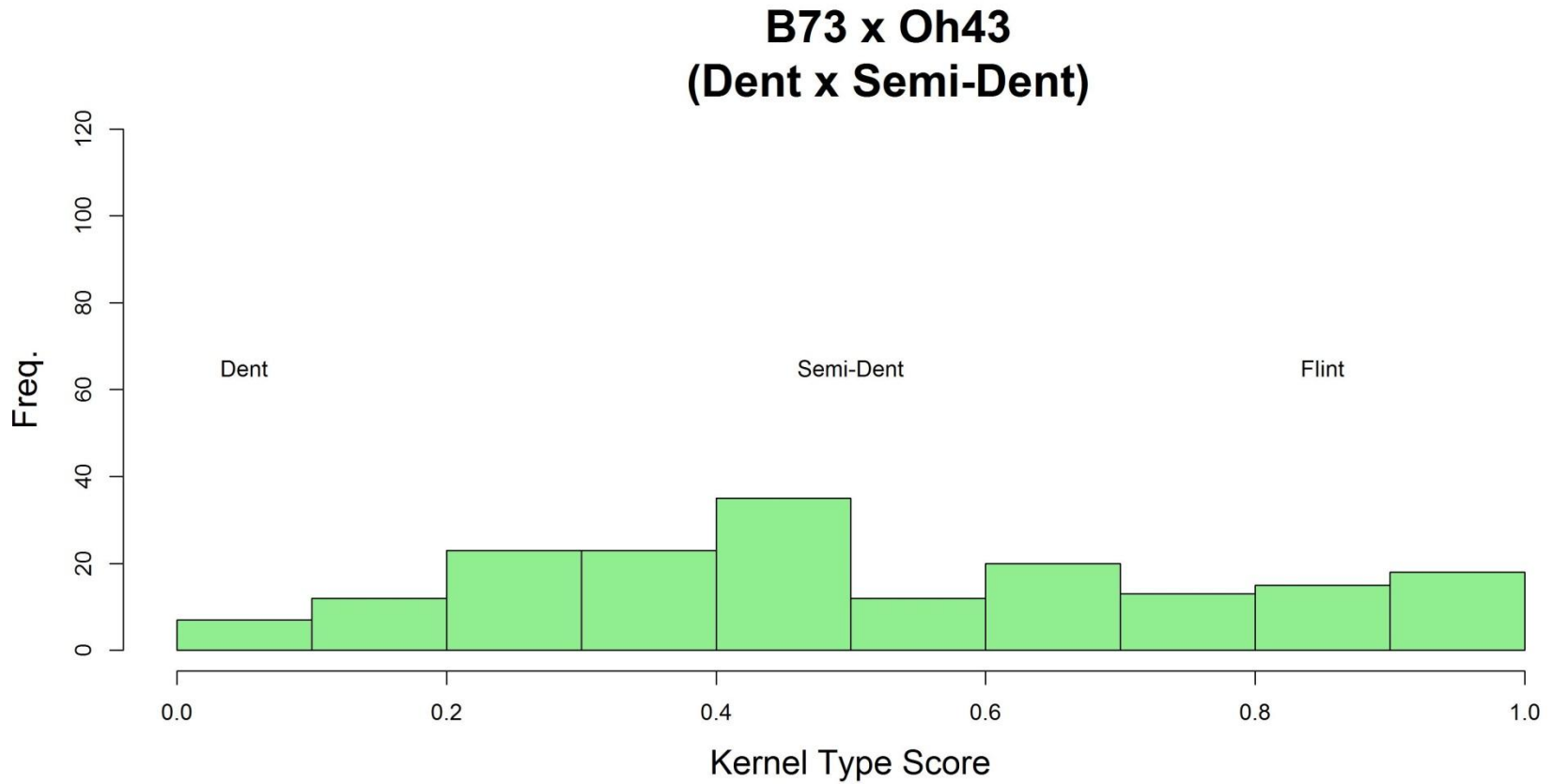Figure S7. Population 9 of NAM, B73 x CML69: A dent x flint cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

B73 x CML69
(Dent x Flint)

Figure S8. Population 12 of NAM, B73 x Ki11: A dent x flint cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

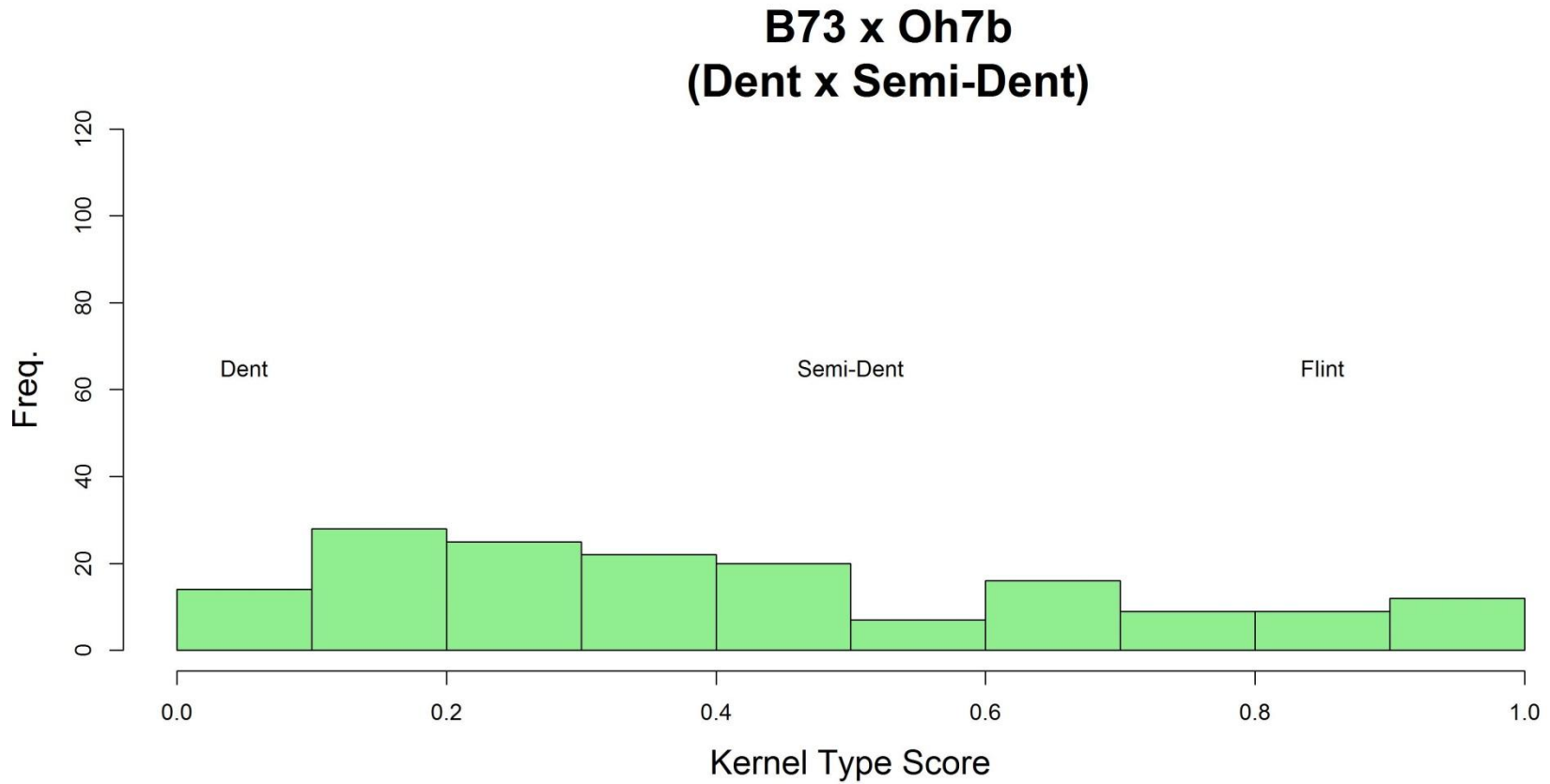**B73 x Ki11**
(Dent x Flint)

Figure S9. Population 13 of NAM, B73 x Ki3: A dent x semi-dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
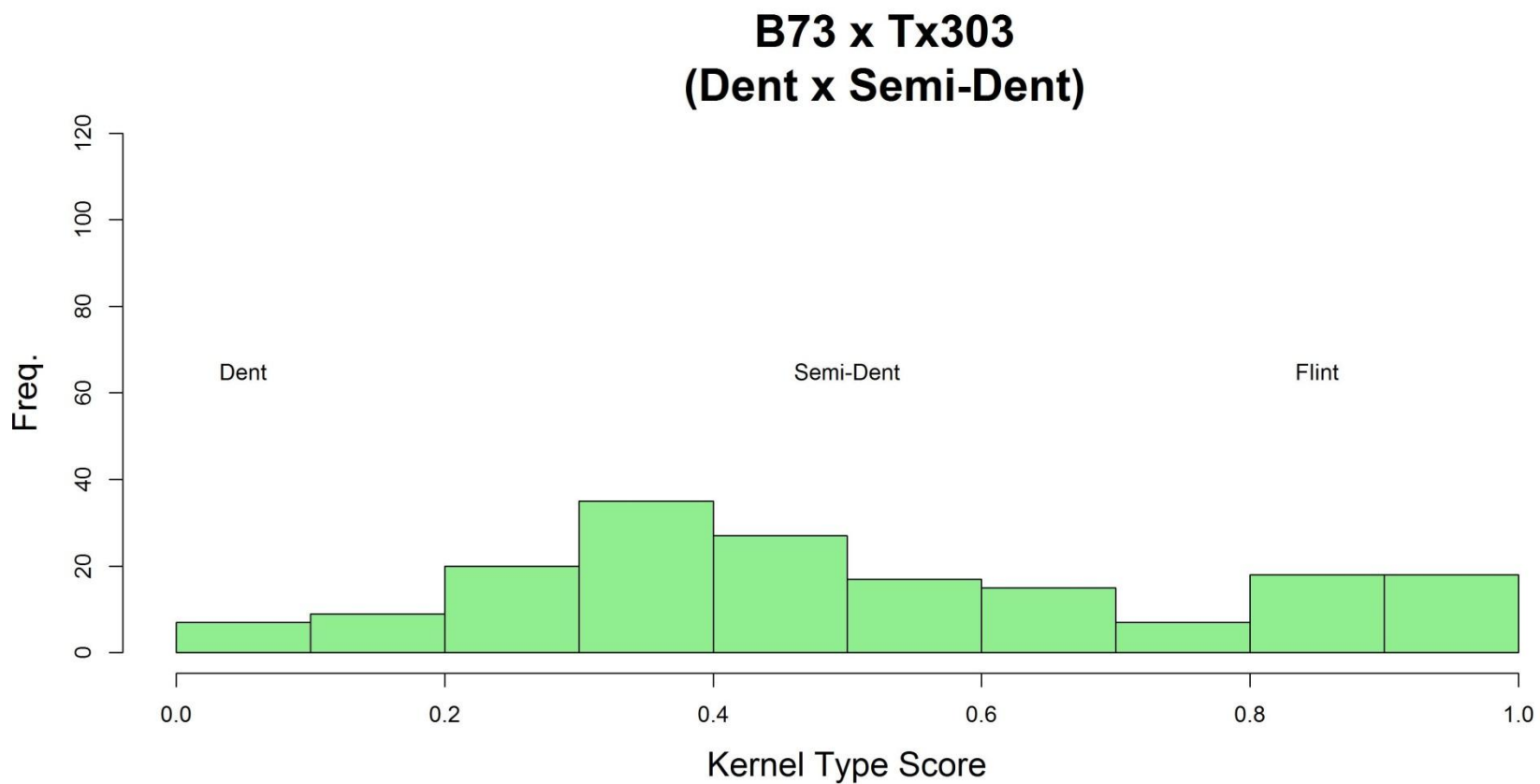

B73 x Ki3
(Dent x Semi-Dent)
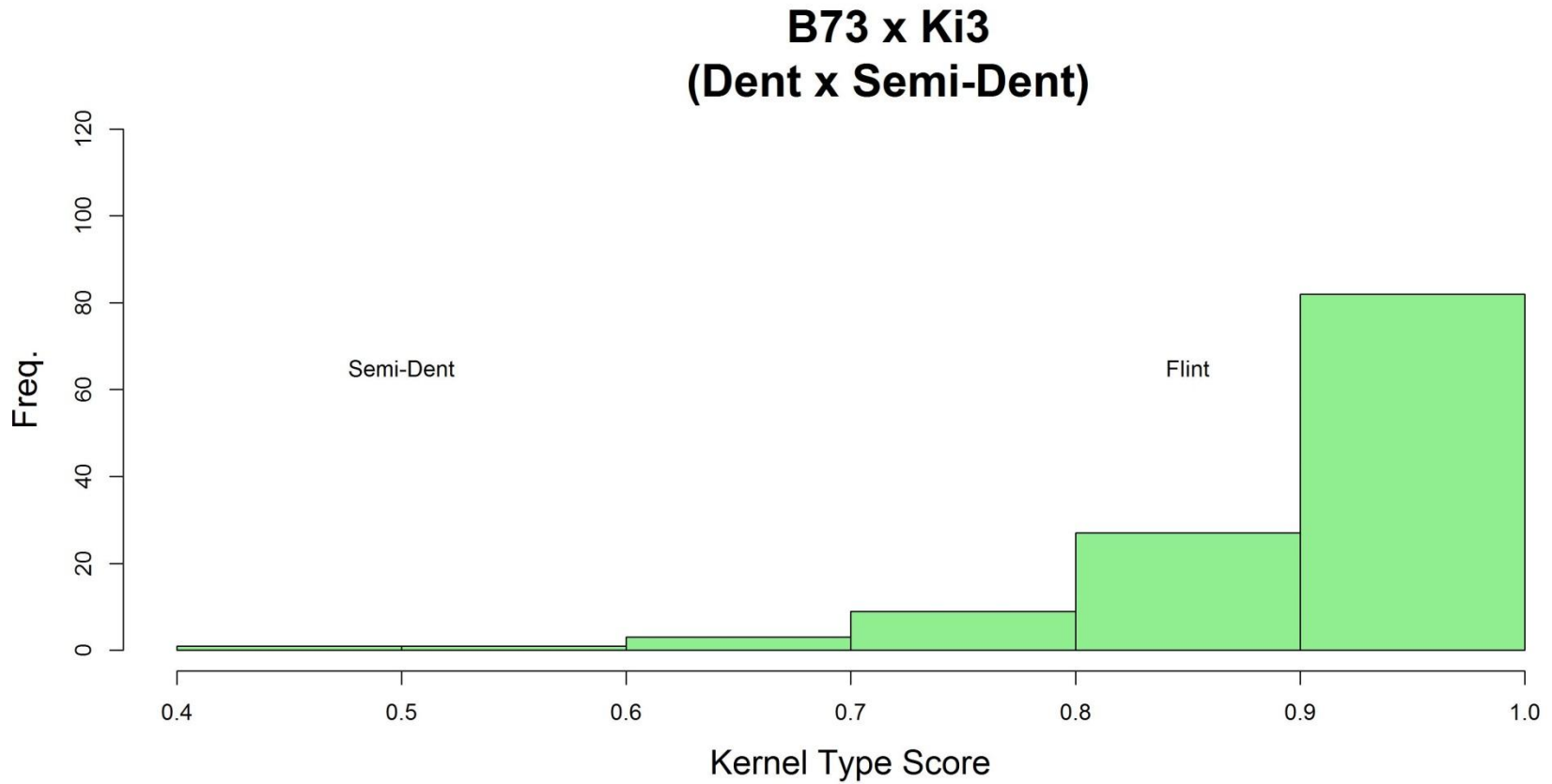
Figure S10. Population 14 of NAM, B73 x Ky21: A dent x dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
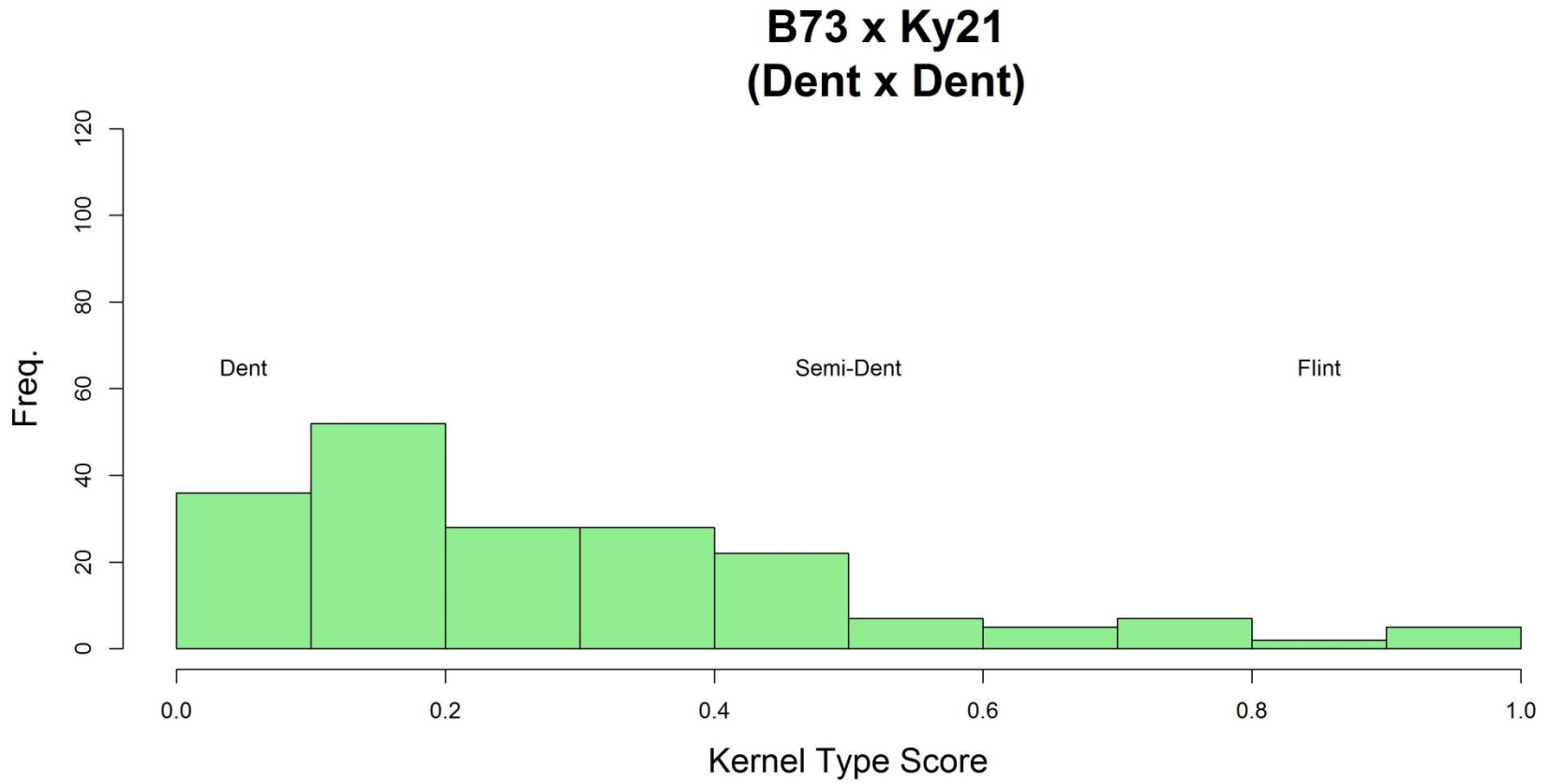
Figure S11. Population 16 of NAM, B73 x M37w: A dent x semi-dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.


**B73 x M37w
(Dent x Semi-Dent)**
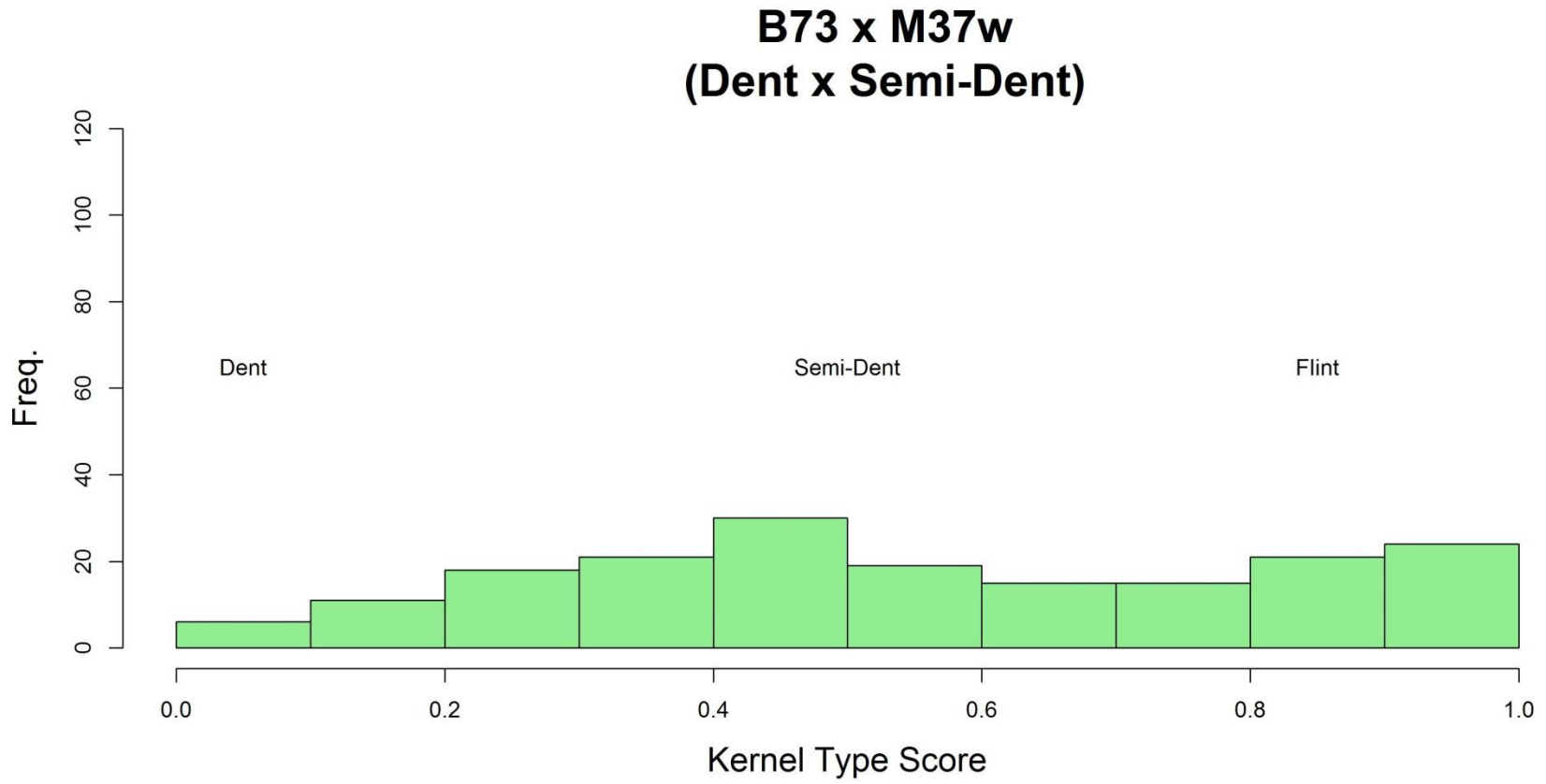
Figure S12. Population 18 of NAM, B73 x CML103: A dent x dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

Figure S13. Population 19 of NAM, B73 x MS71: A dent x dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

Figure S14. Population 20 of NAM, B73 x NC350: A dent x flint cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

Figure S15. Population 21 of NAM, B73 x NC358: A dent x flint cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

**B73 x NC358
(Dent x Flint)**

Figure S16. Population 22 of NAM, B73 x Oh43: A dent x semi-dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

B73 x Oh43
(Dent x Semi-Dent)

Figure S17. Population 23 of NAM, B73 x Oh7b: A dent x semi-dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

Figure S18. Population 25 of NAM, B73 x Tx303: A dent x semi-dent cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.

Figure S19. Population 26 of NAM, B73 x Tzi8: A dent x flint cross. Kernel type score is on the x-axis and frequency is on the y-axis. This is the phenotypic histogram from the RILs in this cross.
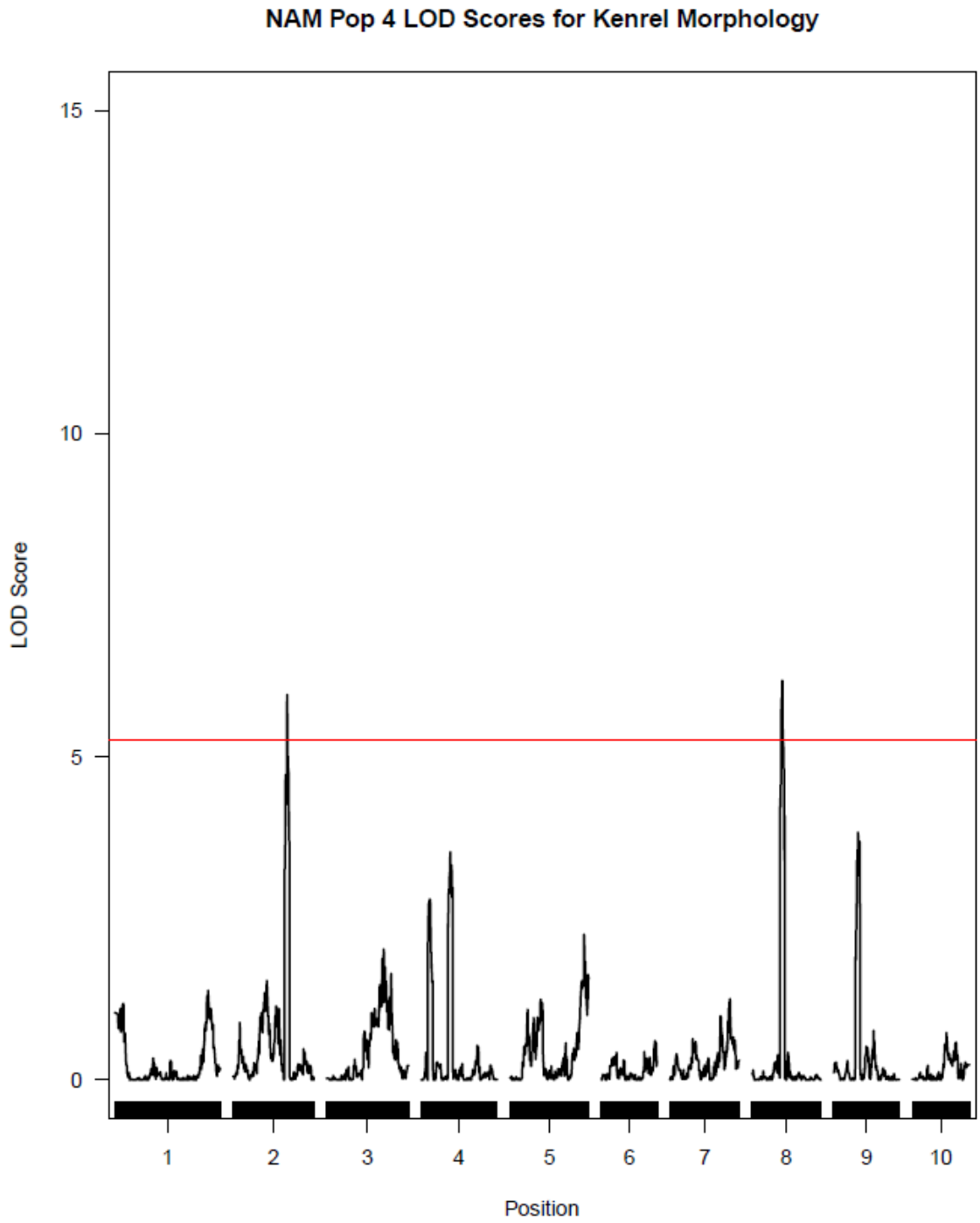

B73 x Tzi8
(Dent x Flint)
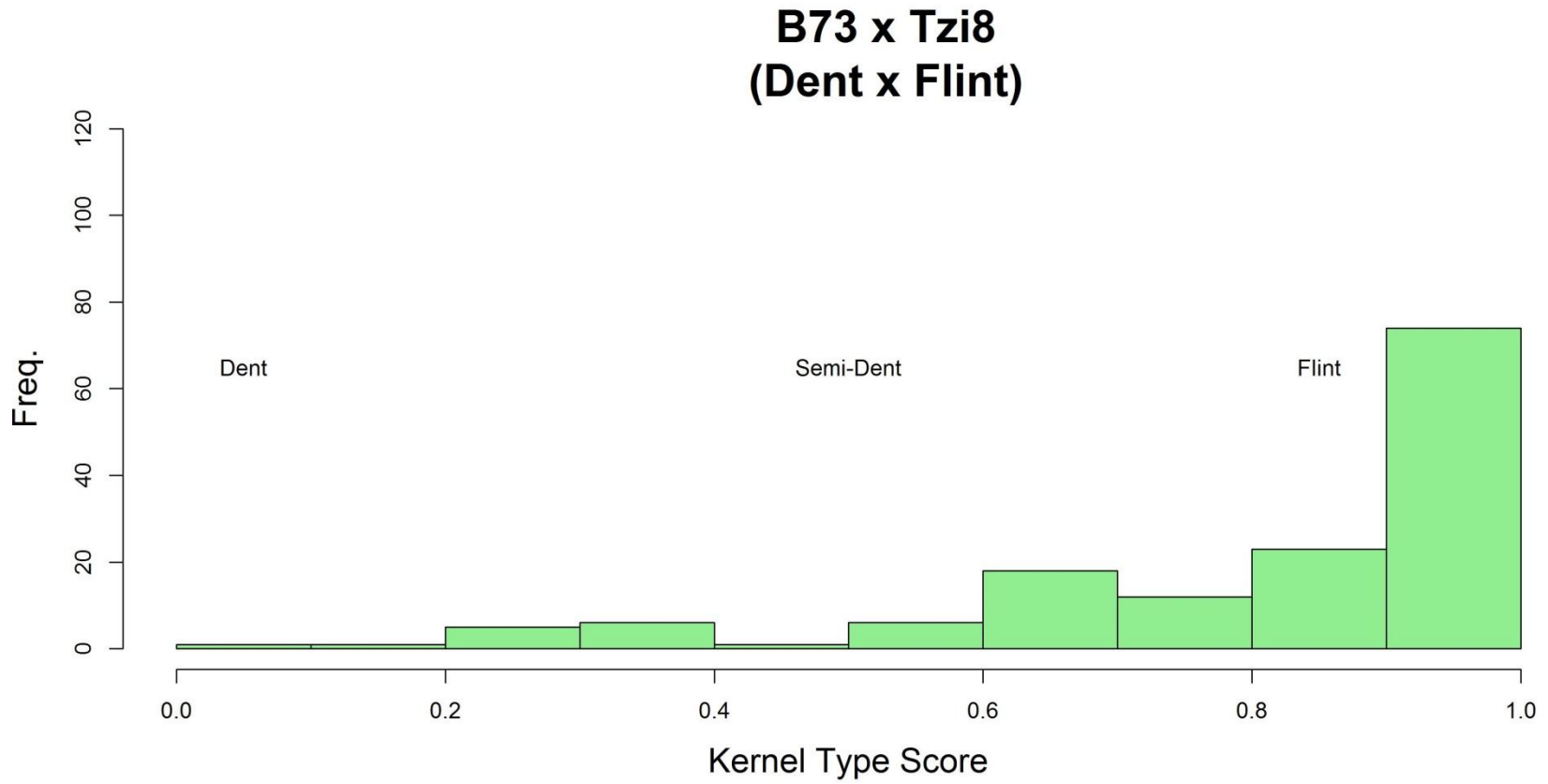
Figure S20. NAM population 1 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
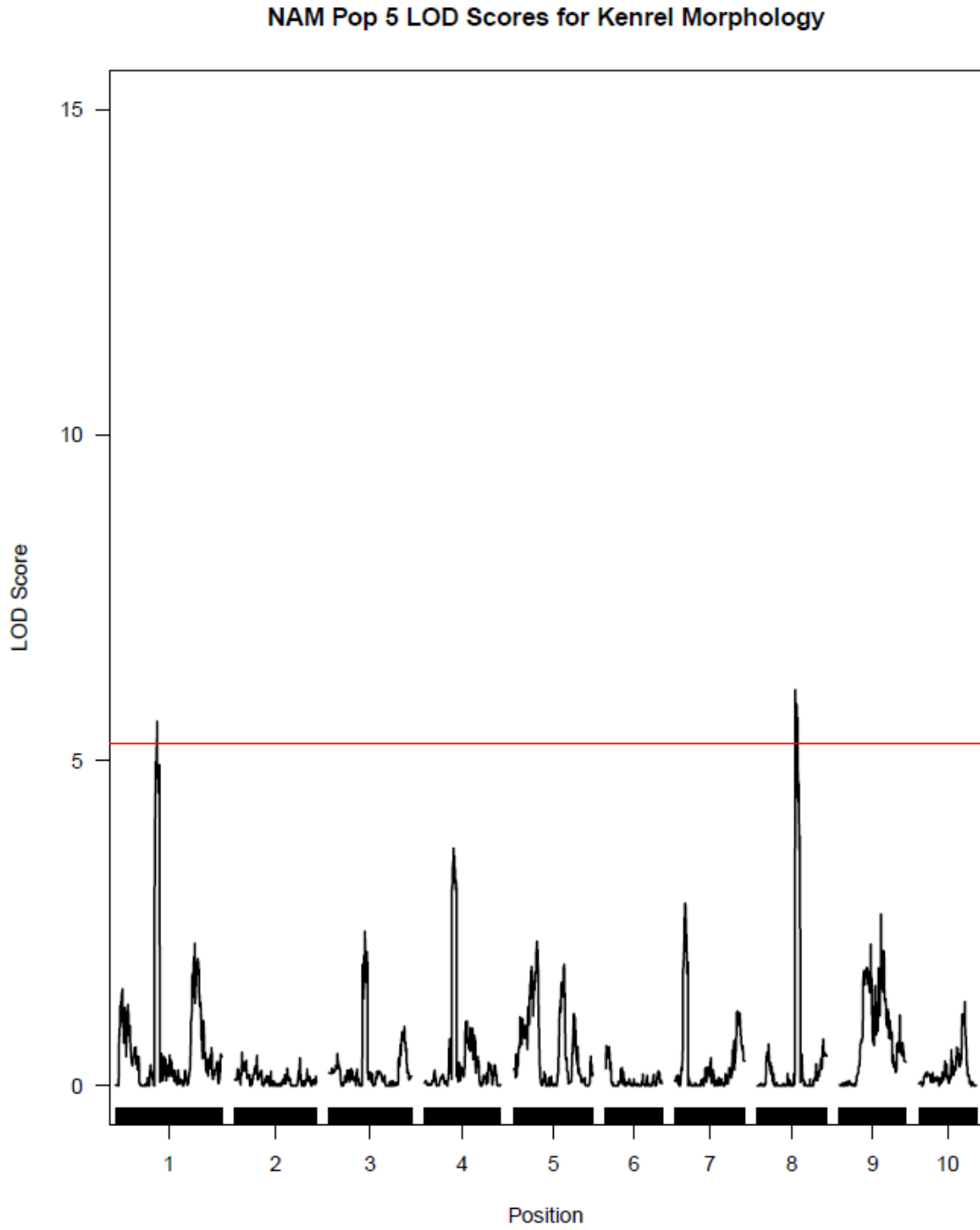

NAM Pop 1 LOD Scores for Kernel Morphology
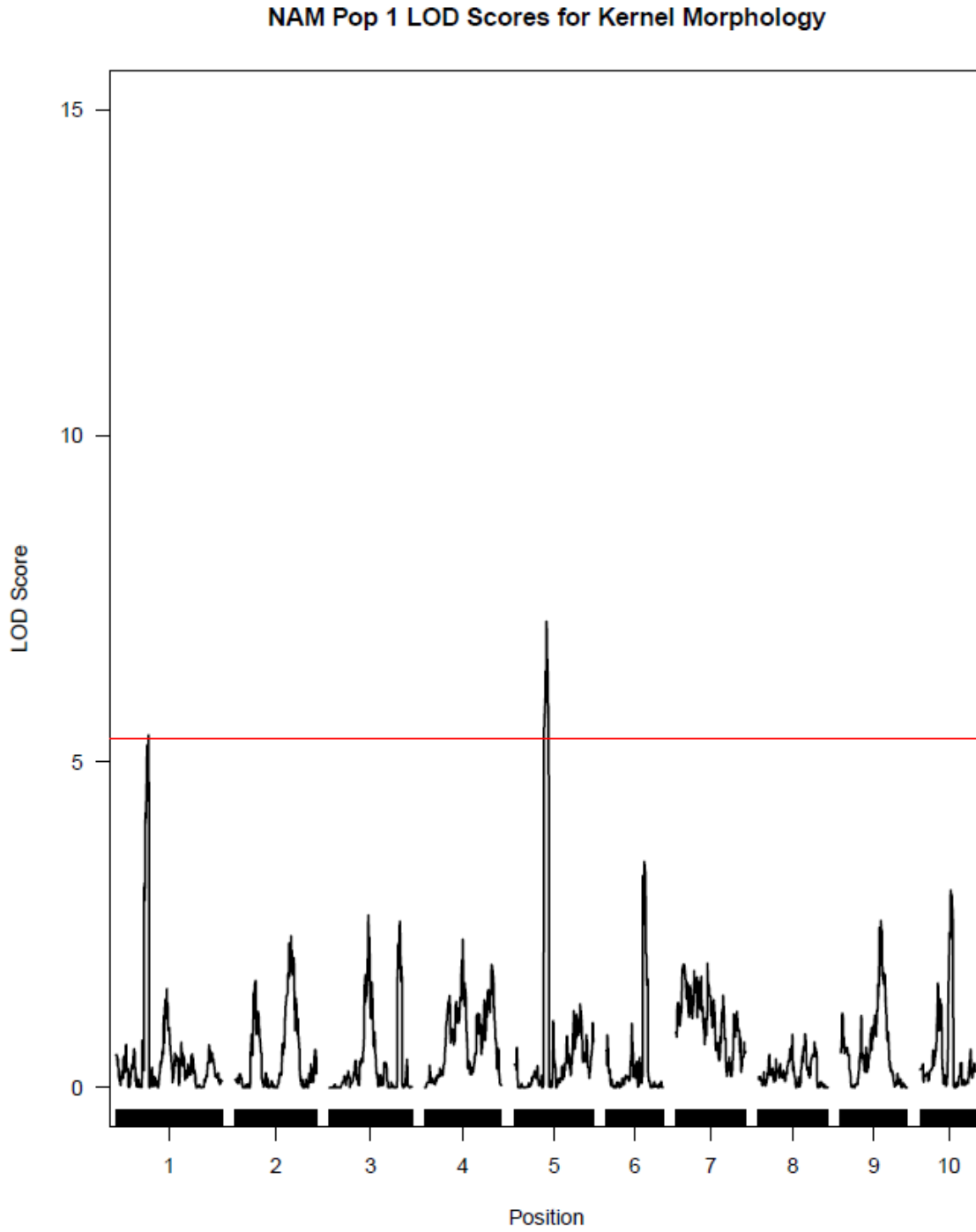
Figure S21. NAM population 2 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
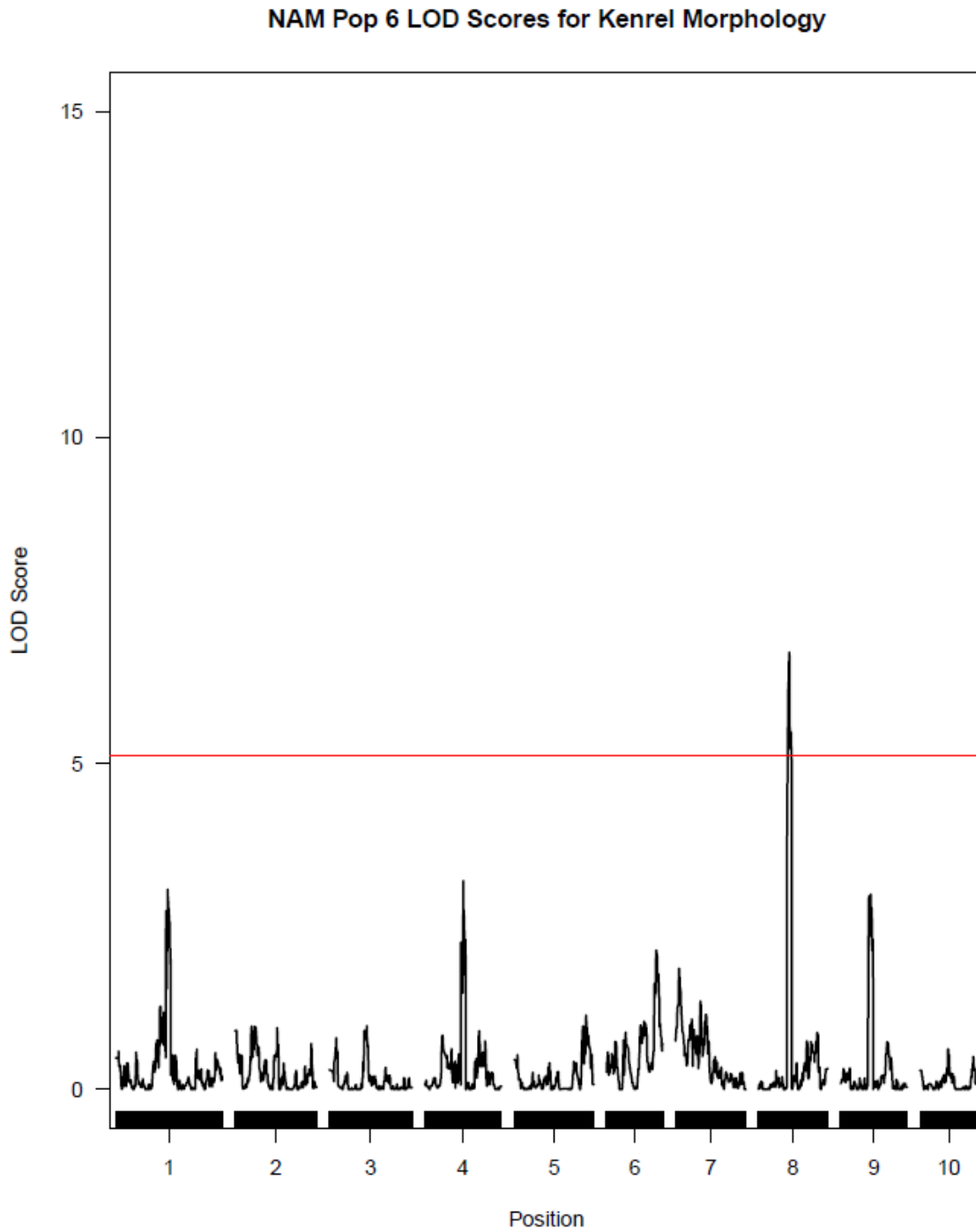

NAM Pop 2 LOD Scores for Kenrel Morphology
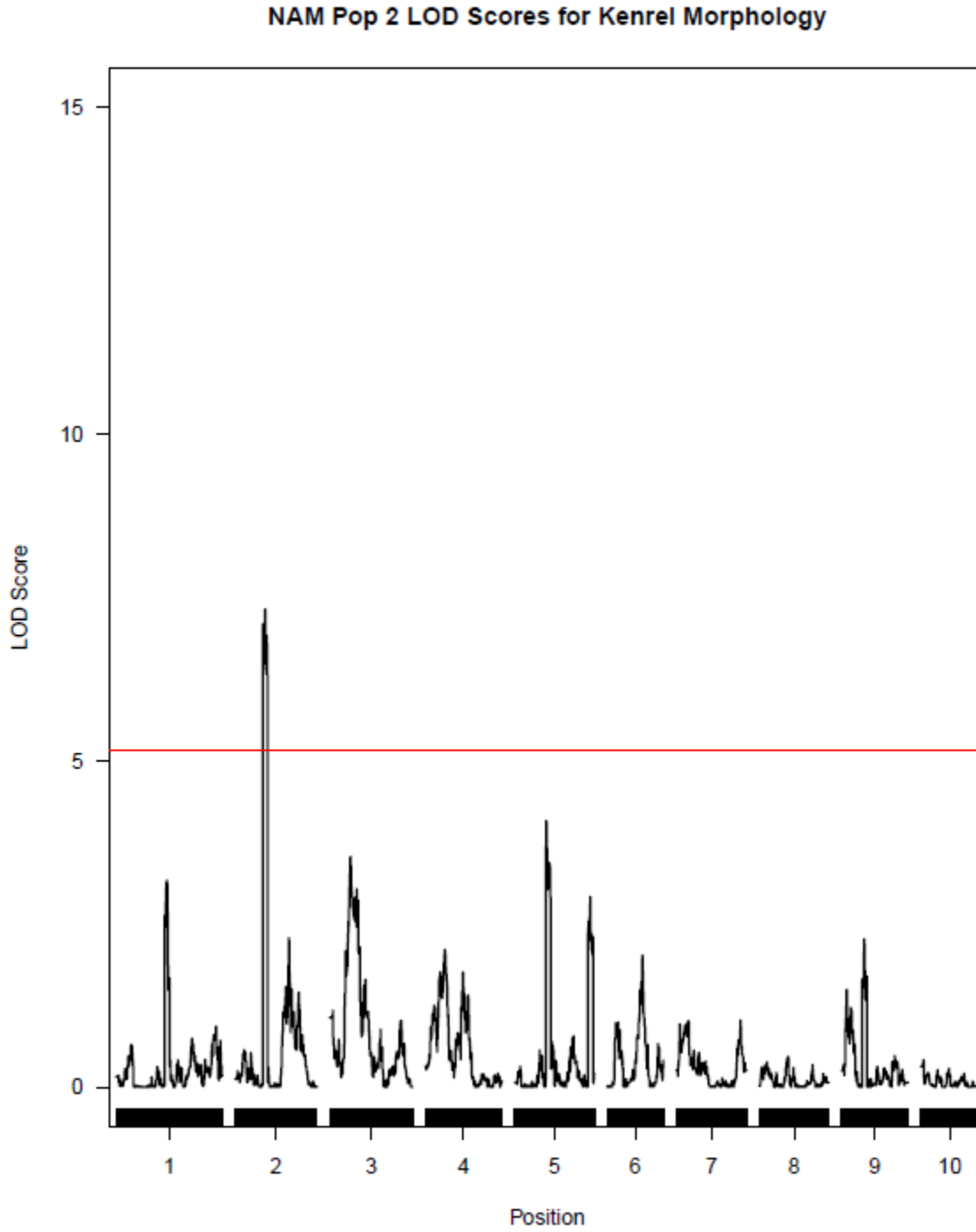
Figure S22. NAM population 3 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.

NAM Pop 3 LOD Scores for Kenrel Morphology
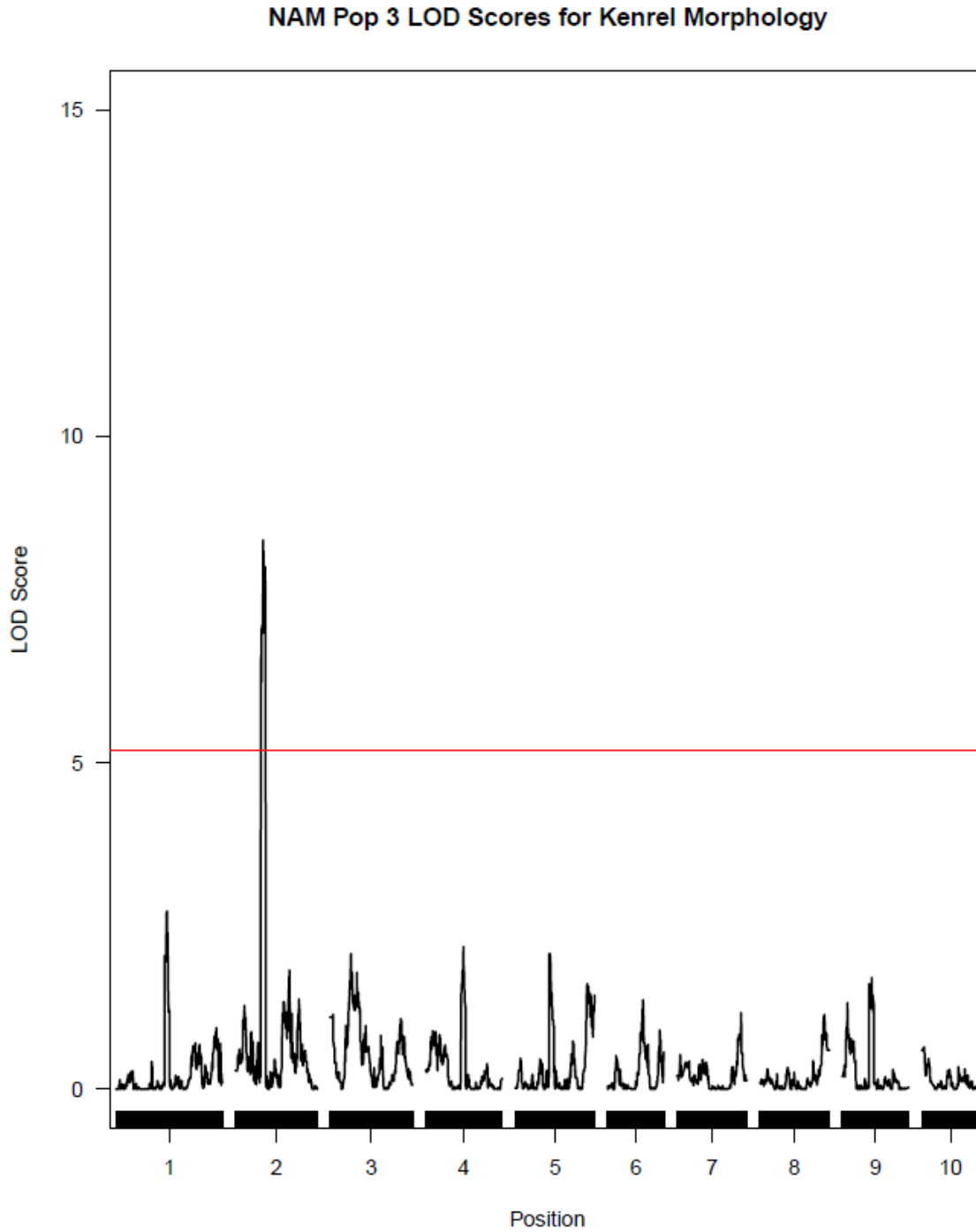
Figure S23. NAM population 4 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
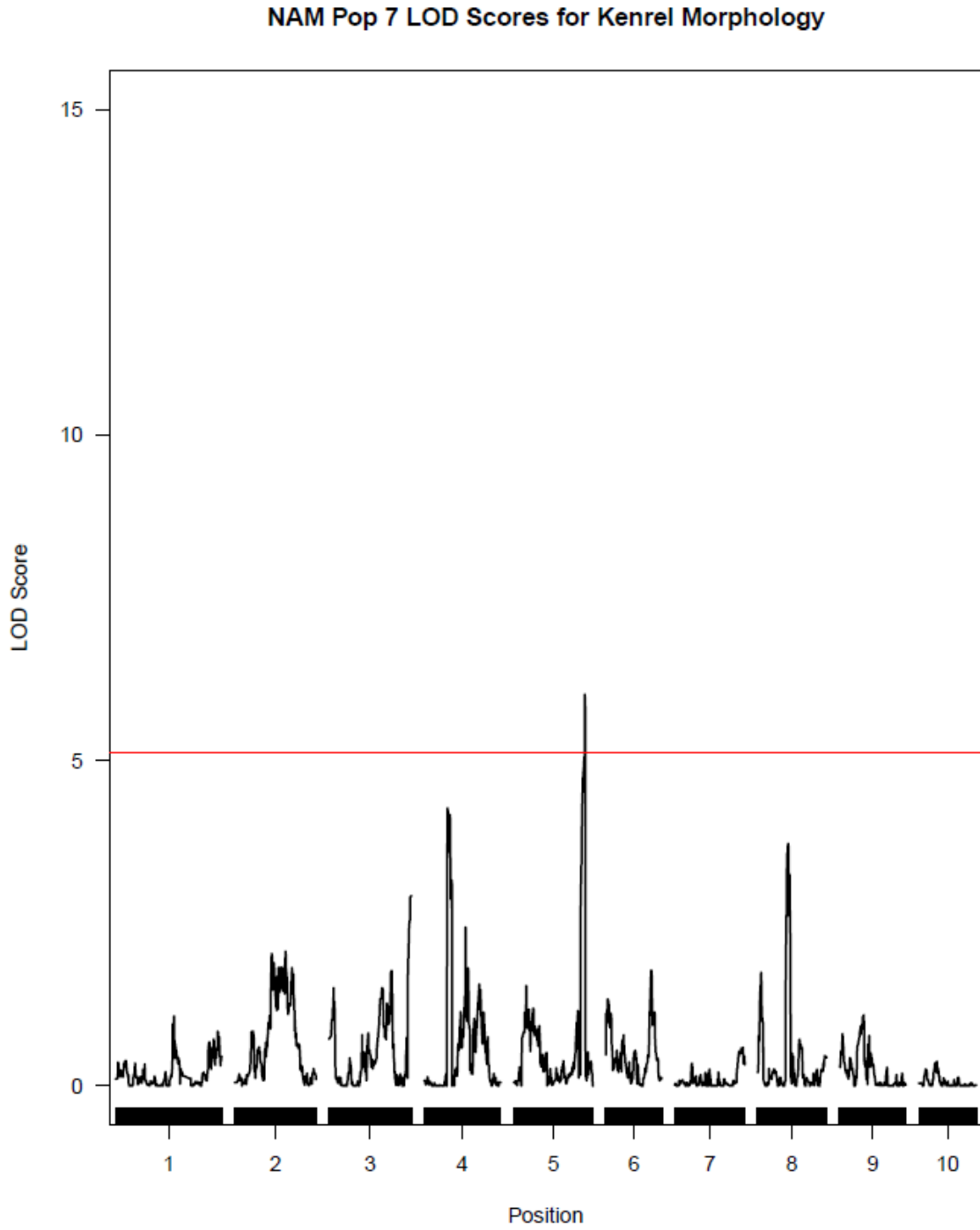

NAM Pop 4 LOD Scores for Kenrel Morphology

Figure S24. NAM population 5 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
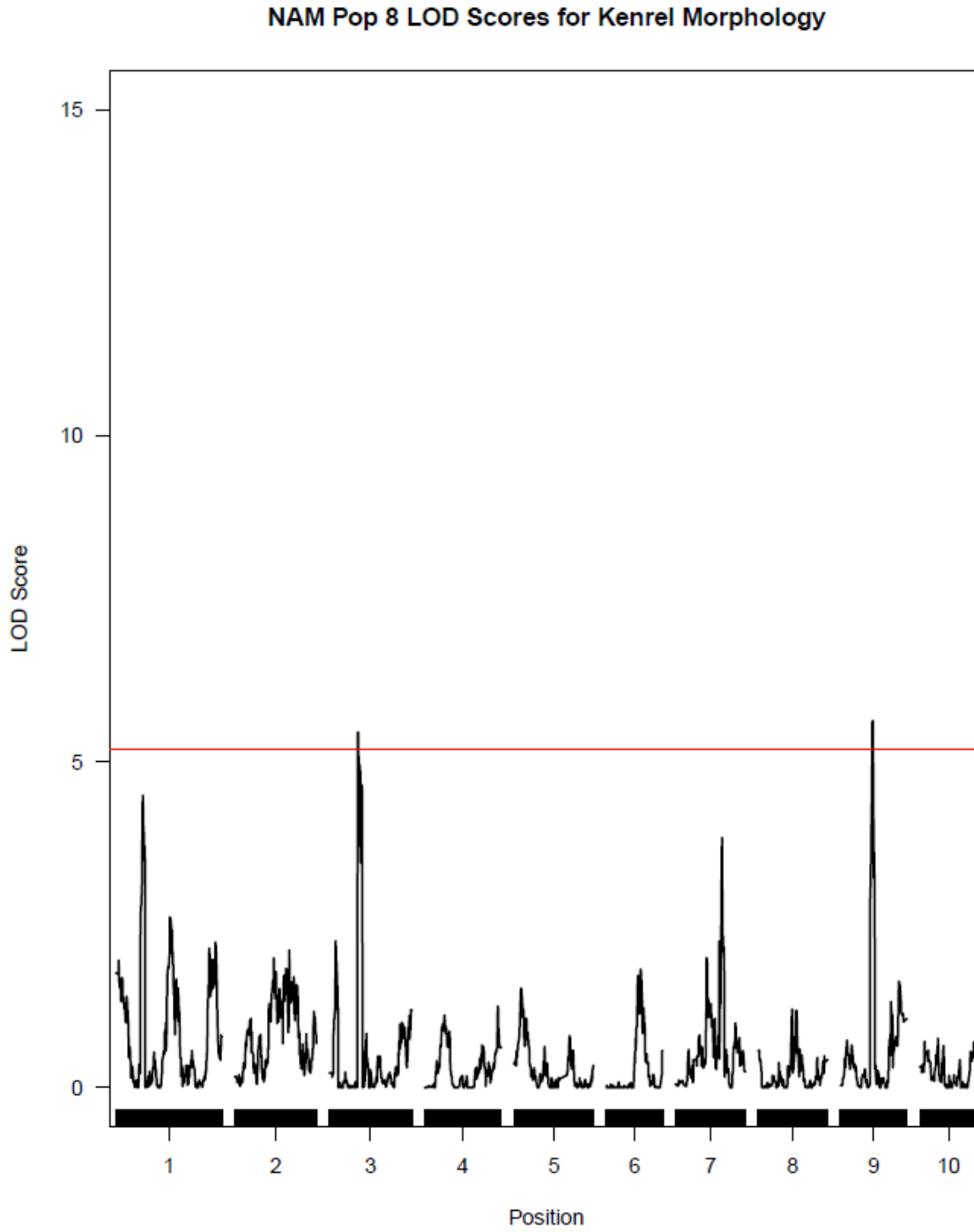

NAM Pop 5 LOD Scores for Kenrel Morphology

Figure S25. NAM population 6 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
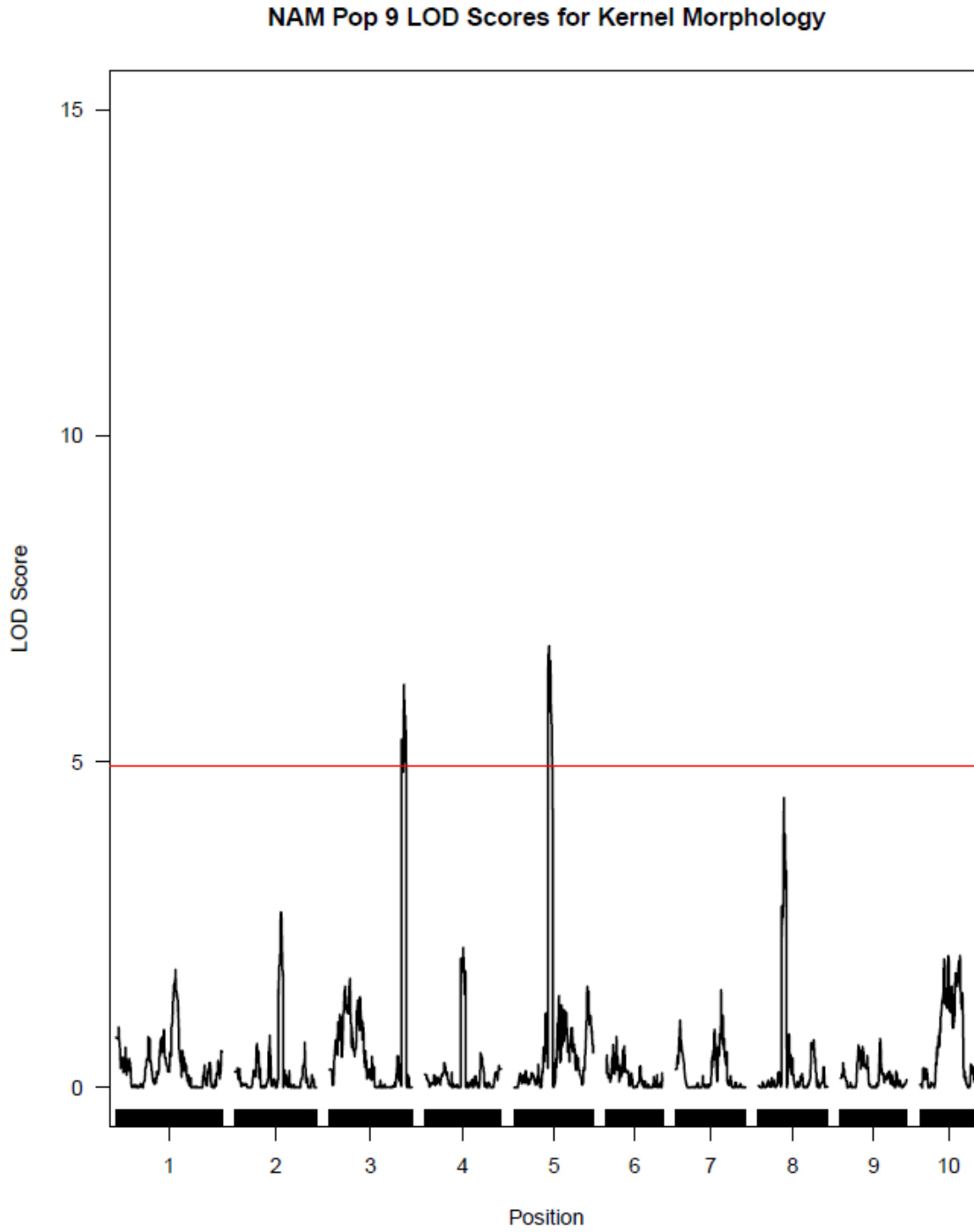

NAM Pop 6 LOD Scores for Kenrel Morphology

Figure S26. NAM population 7 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
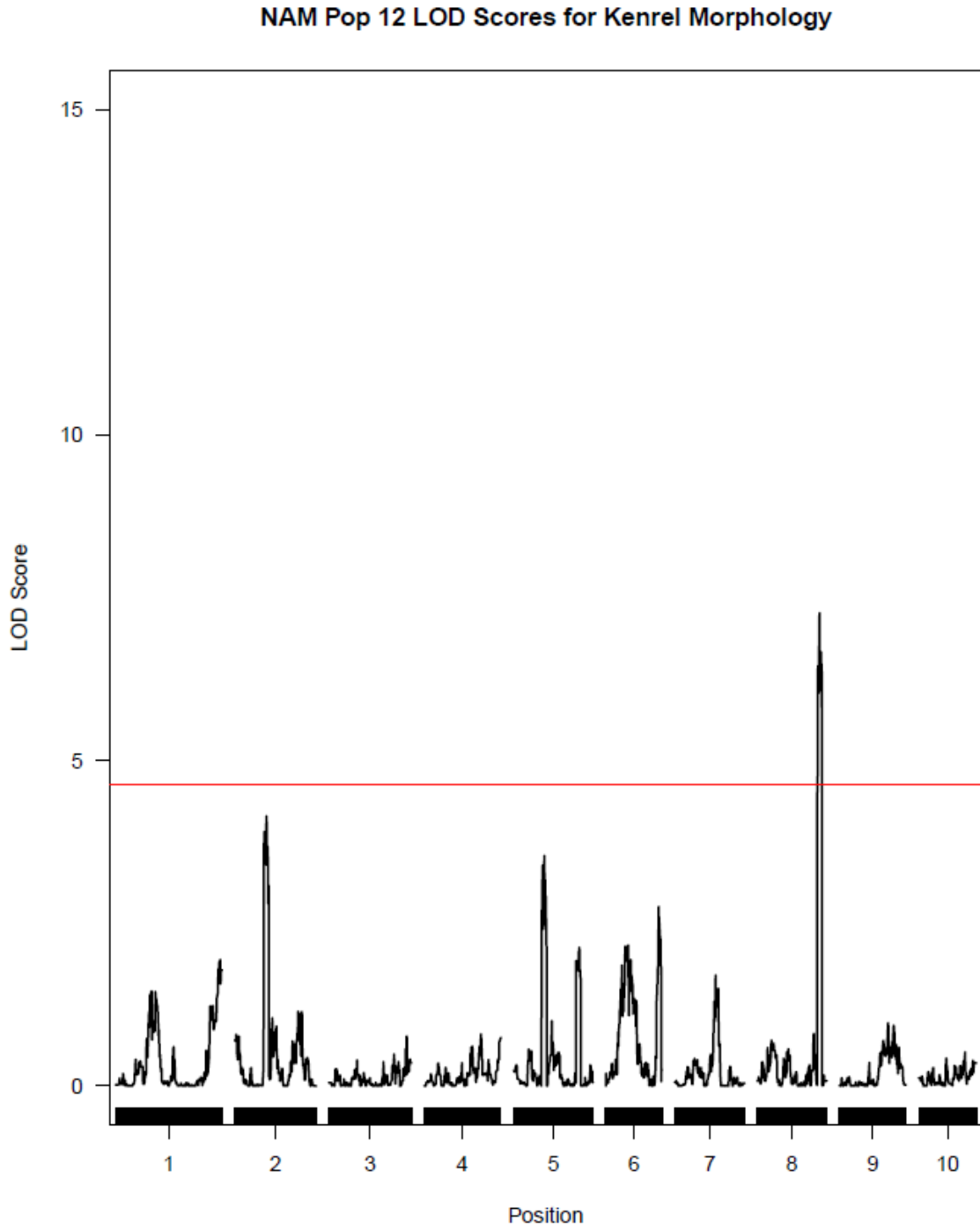

NAM Pop 7 LOD Scores for Kenrel Morphology

Figure S27. NAM population 8 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.

NAM Pop 8 LOD Scores for Kenrel Morphology

Figure S28. NAM population 9 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
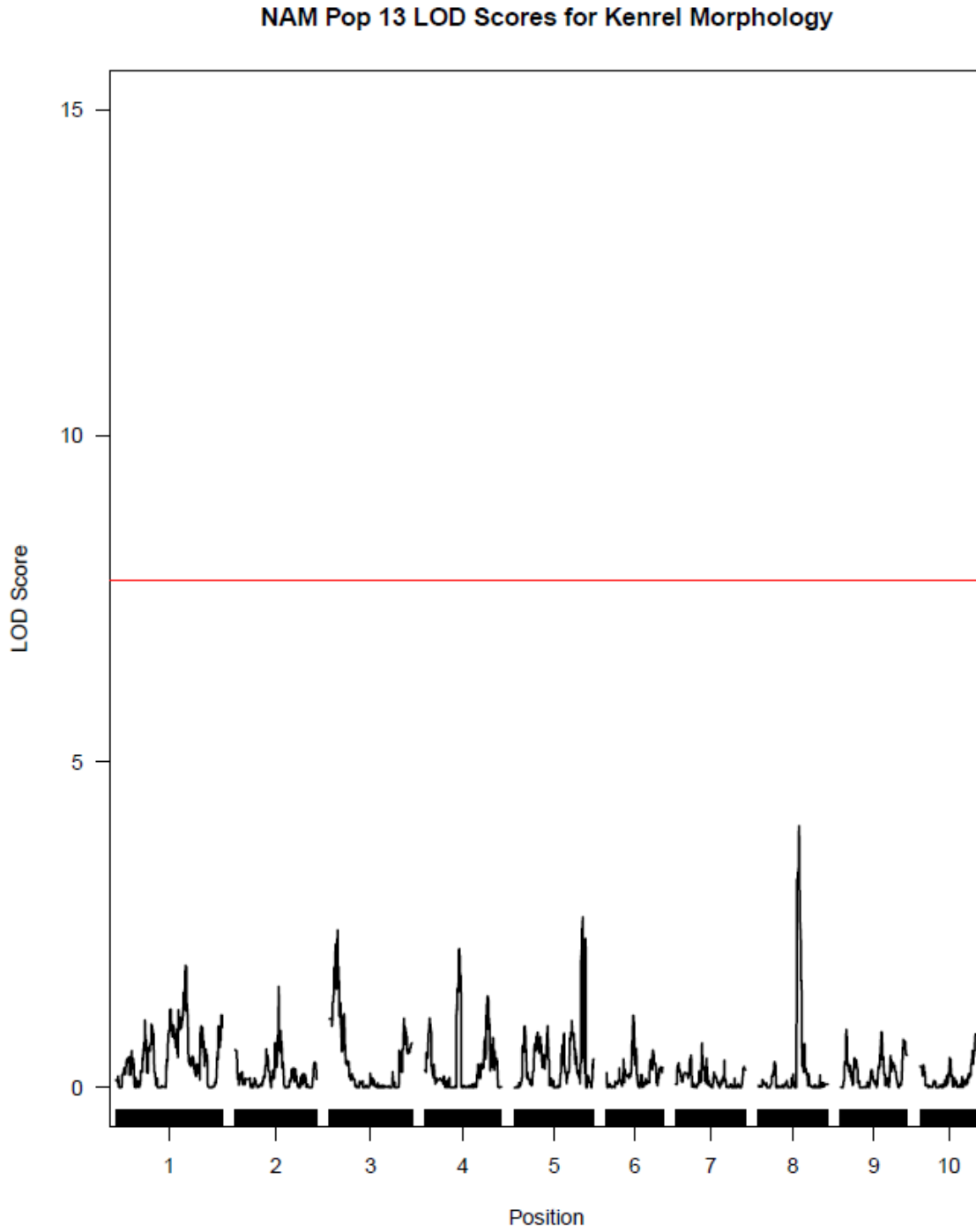

NAM Pop 9 LOD Scores for Kernel Morphology

Figure S29. NAM population 12 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
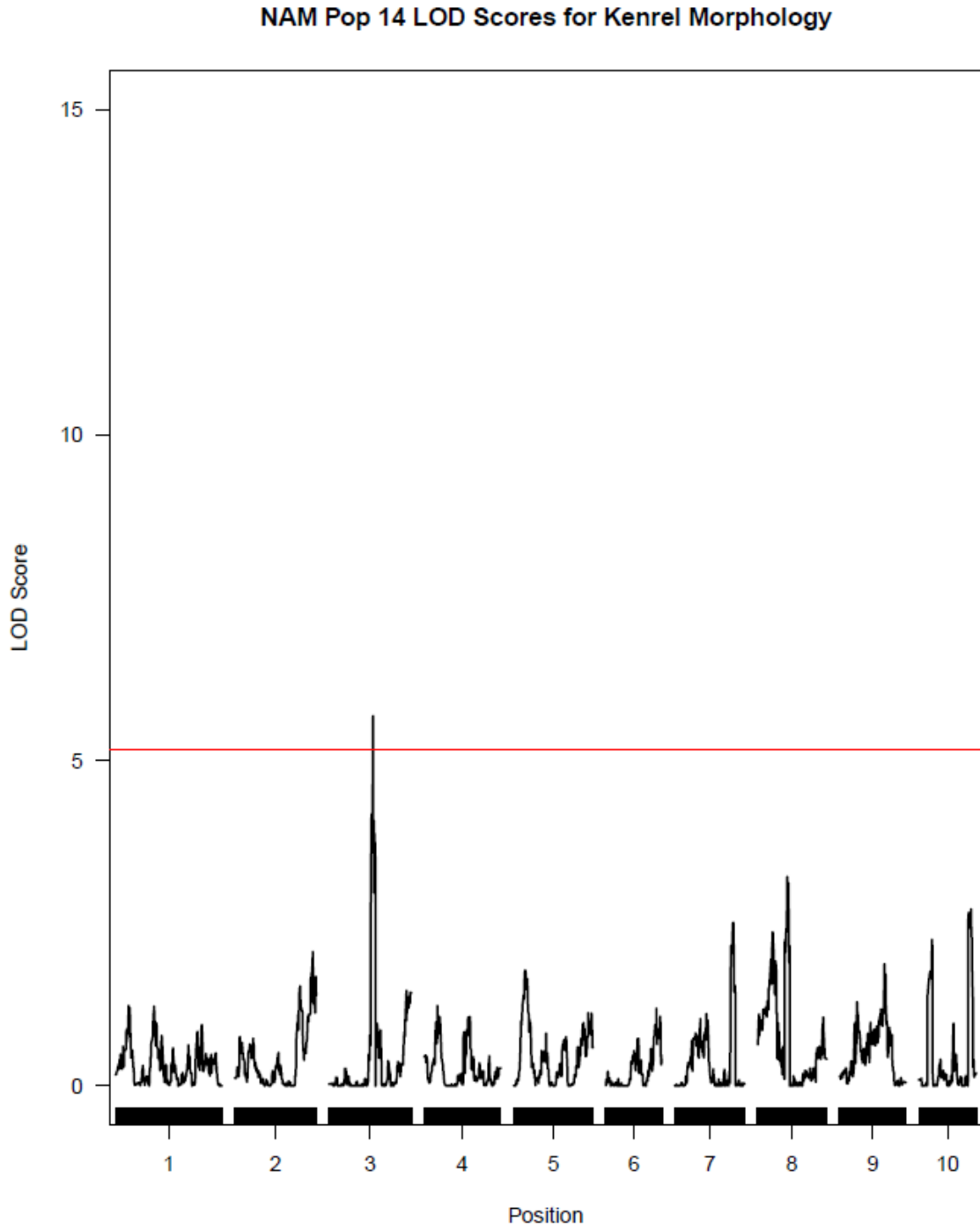

NAM Pop 12 LOD Scores for Kenrel Morphology

Figure S30. NAM population 13 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
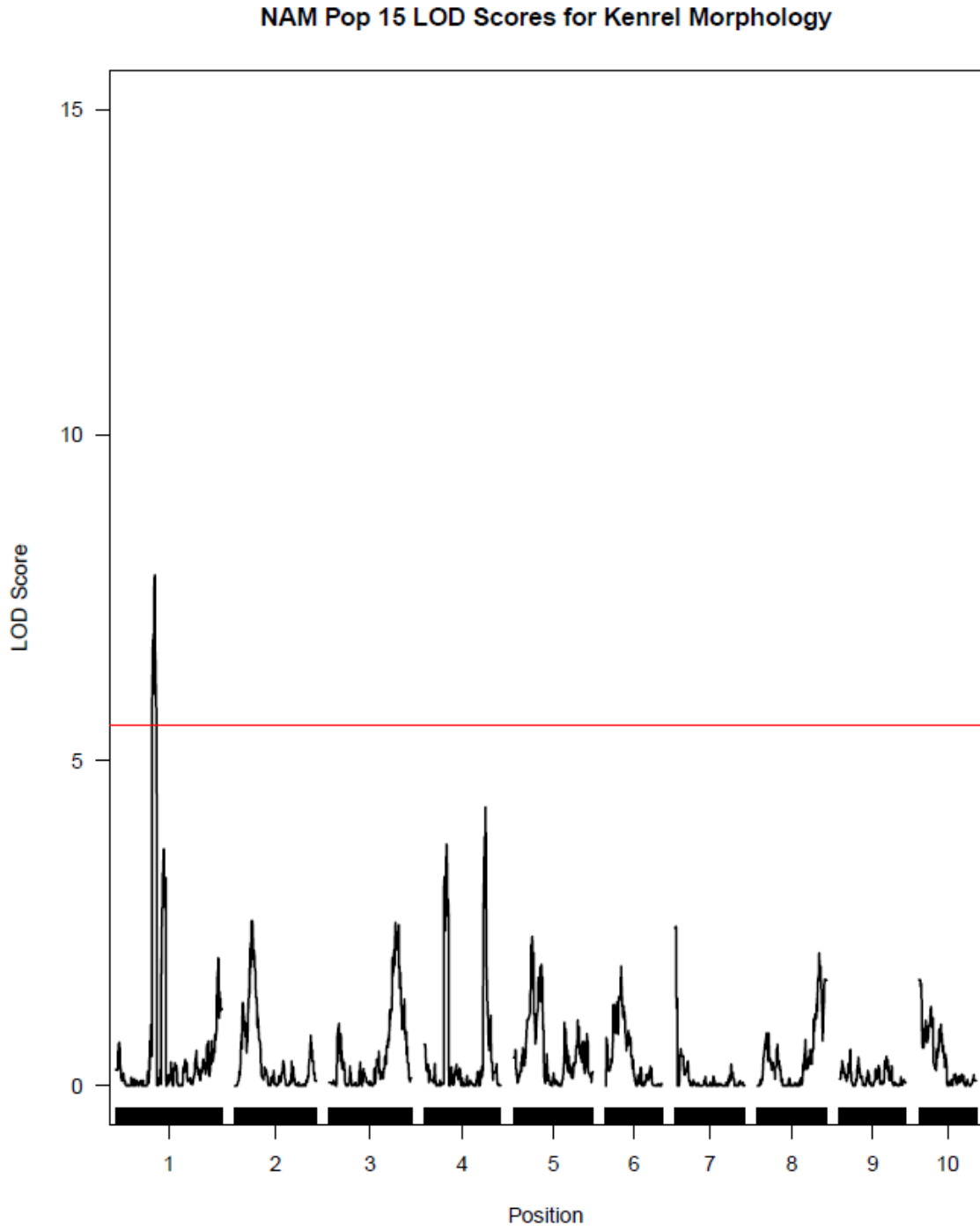

NAM Pop 13 LOD Scores for Kenrel Morphology

Figure S31. NAM population 14 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
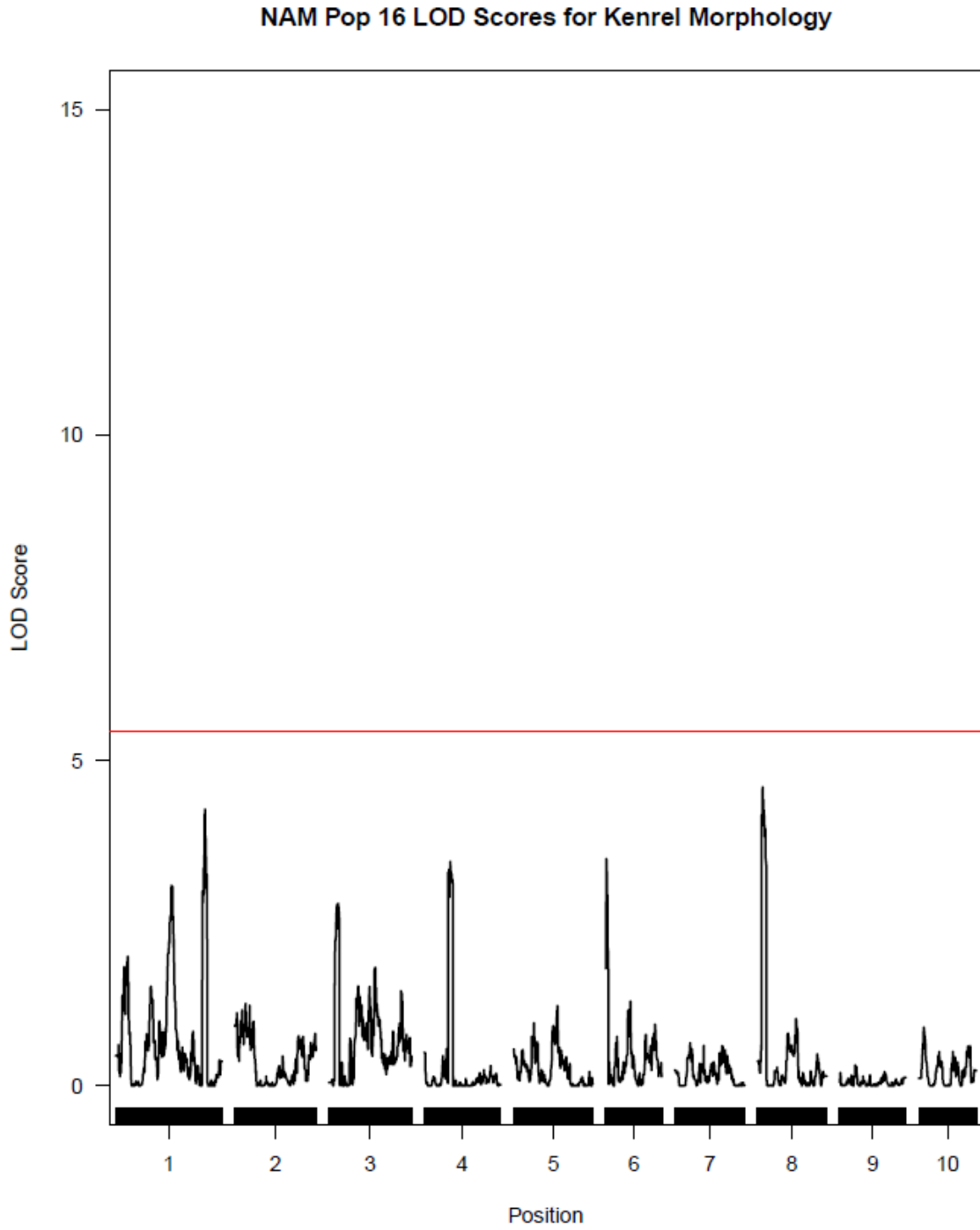

NAM Pop 14 LOD Scores for Kenrel Morphology

Figure S32. NAM population 15 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.

NAM Pop 15 LOD Scores for Kenrel Morphology

Figure S33. NAM population 16 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
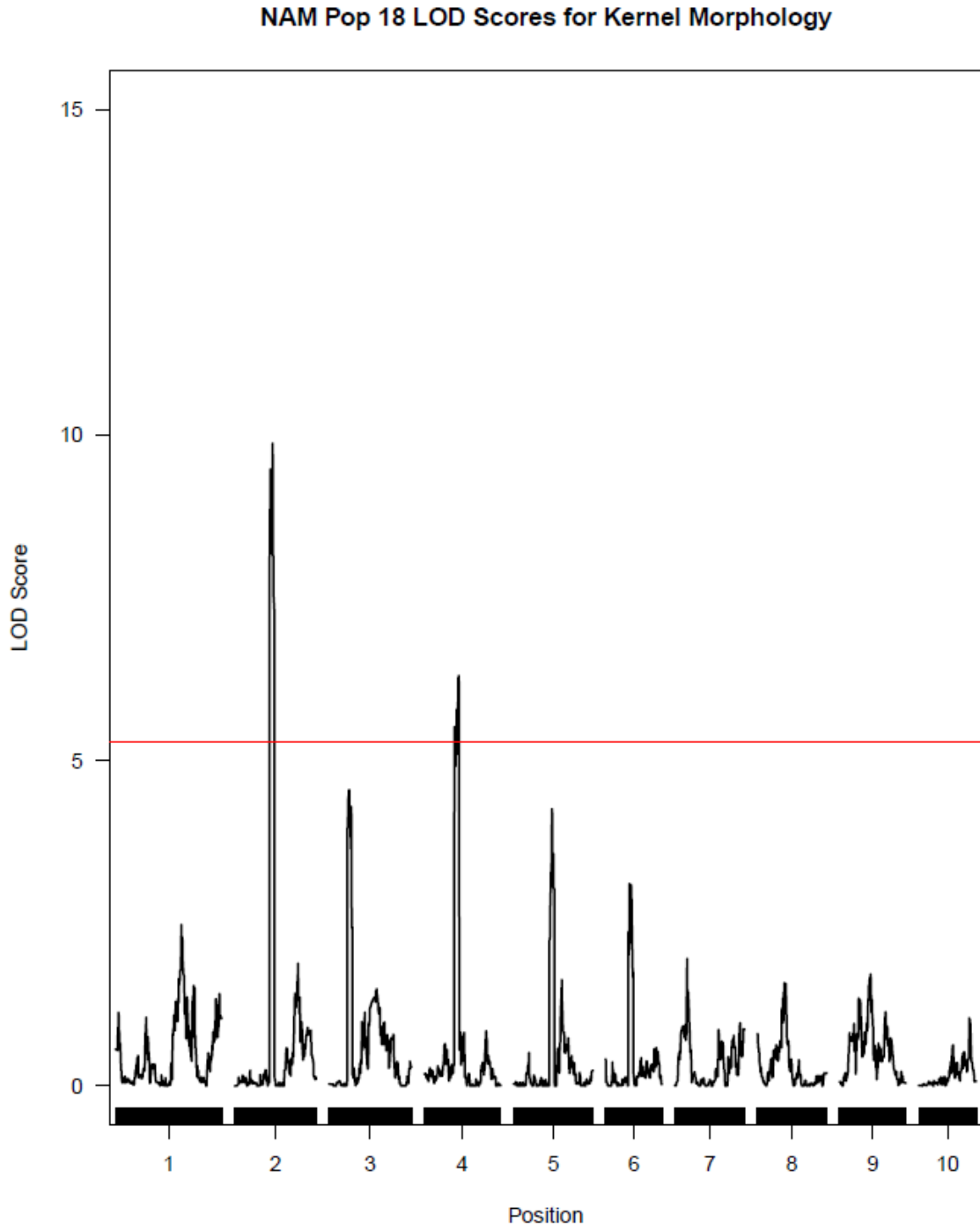

NAM Pop 16 LOD Scores for Kenrel Morphology

Figure S34. NAM population 18 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
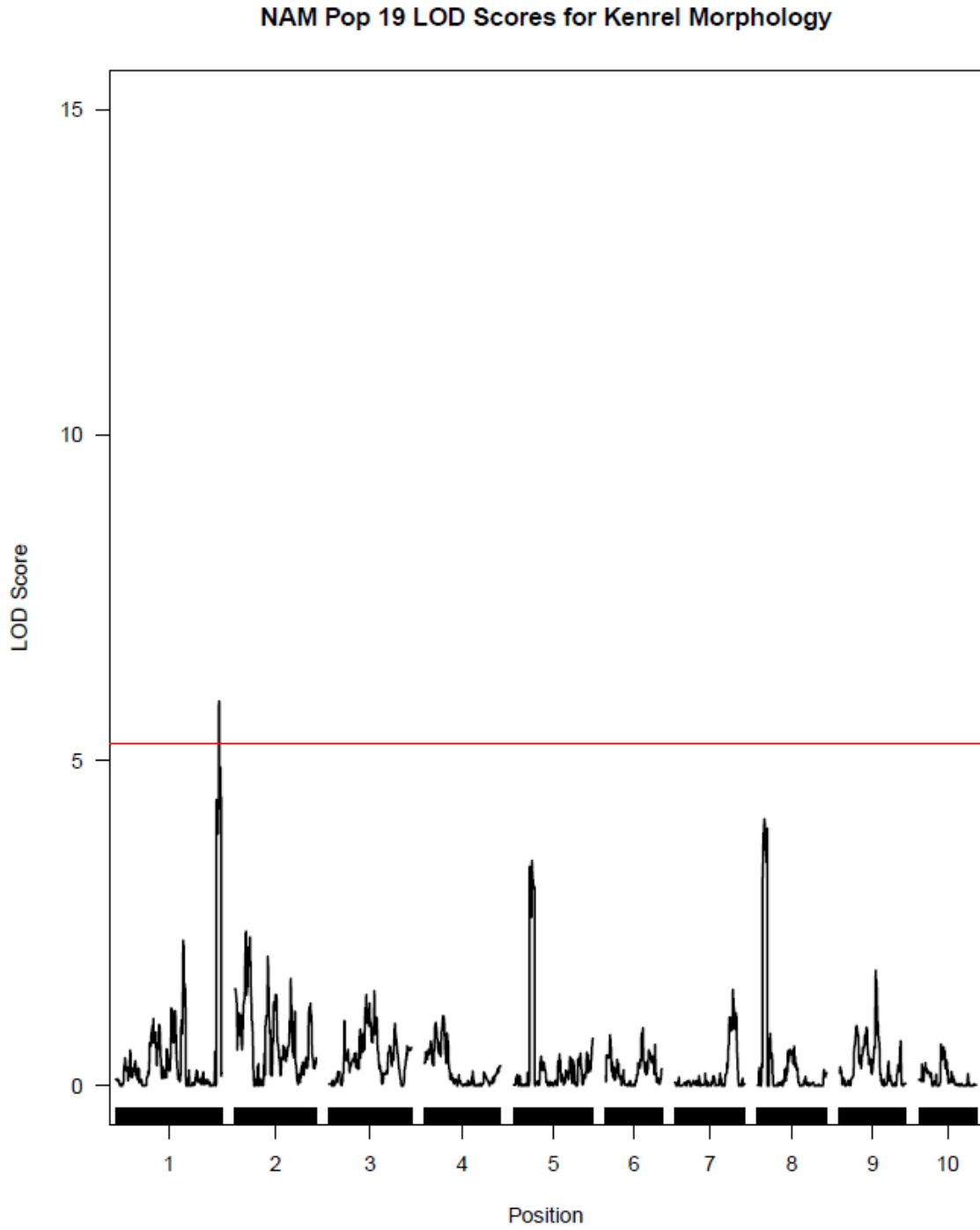

NAM Pop 18 LOD Scores for Kernel Morphology

Figure S35. NAM population 19 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
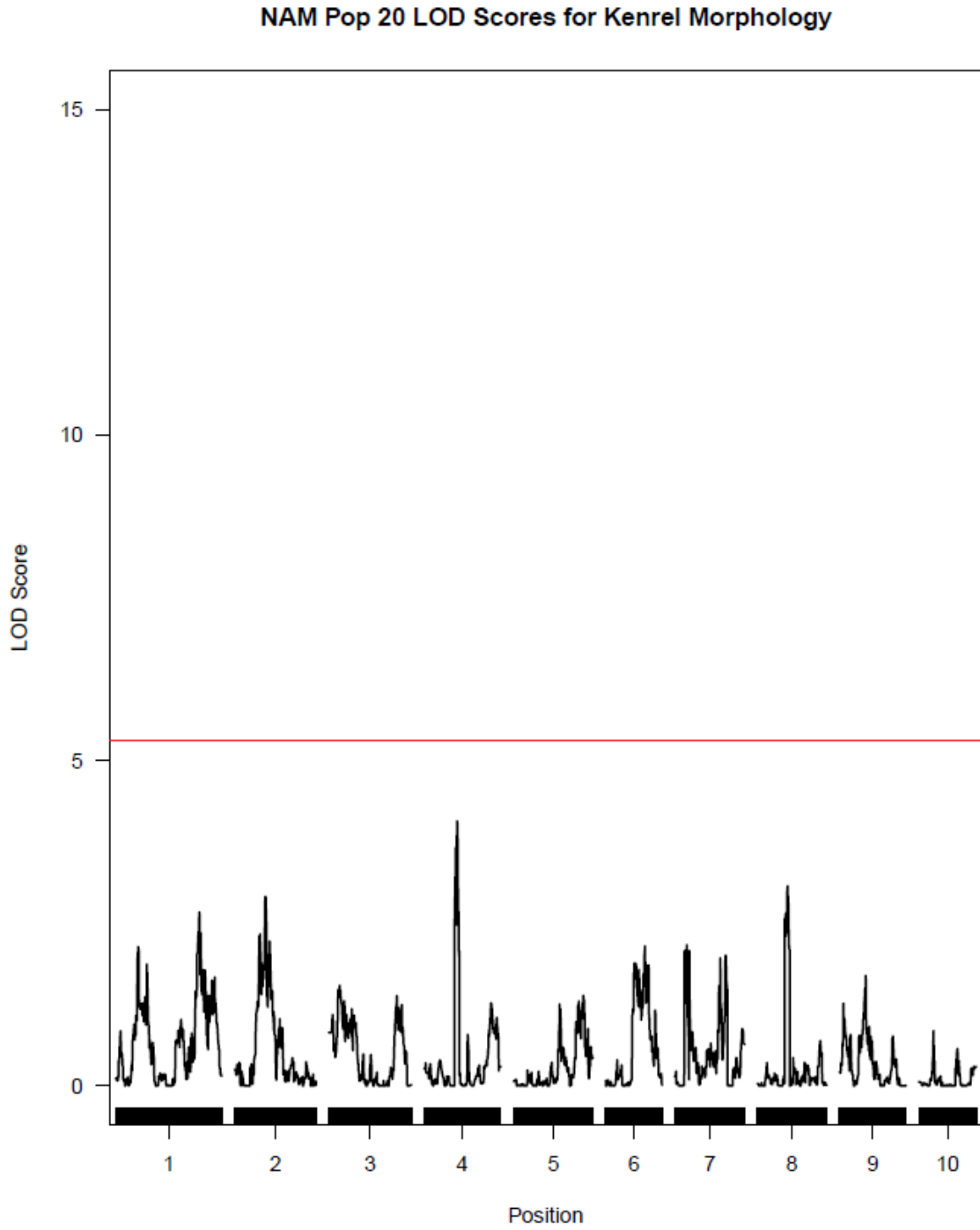

NAM Pop 19 LOD Scores for Kenrel Morphology

Figure S36. NAM population 20 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.

NAM Pop 20 LOD Scores for Kenrel Morphology

Figure S37. NAM population 21 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
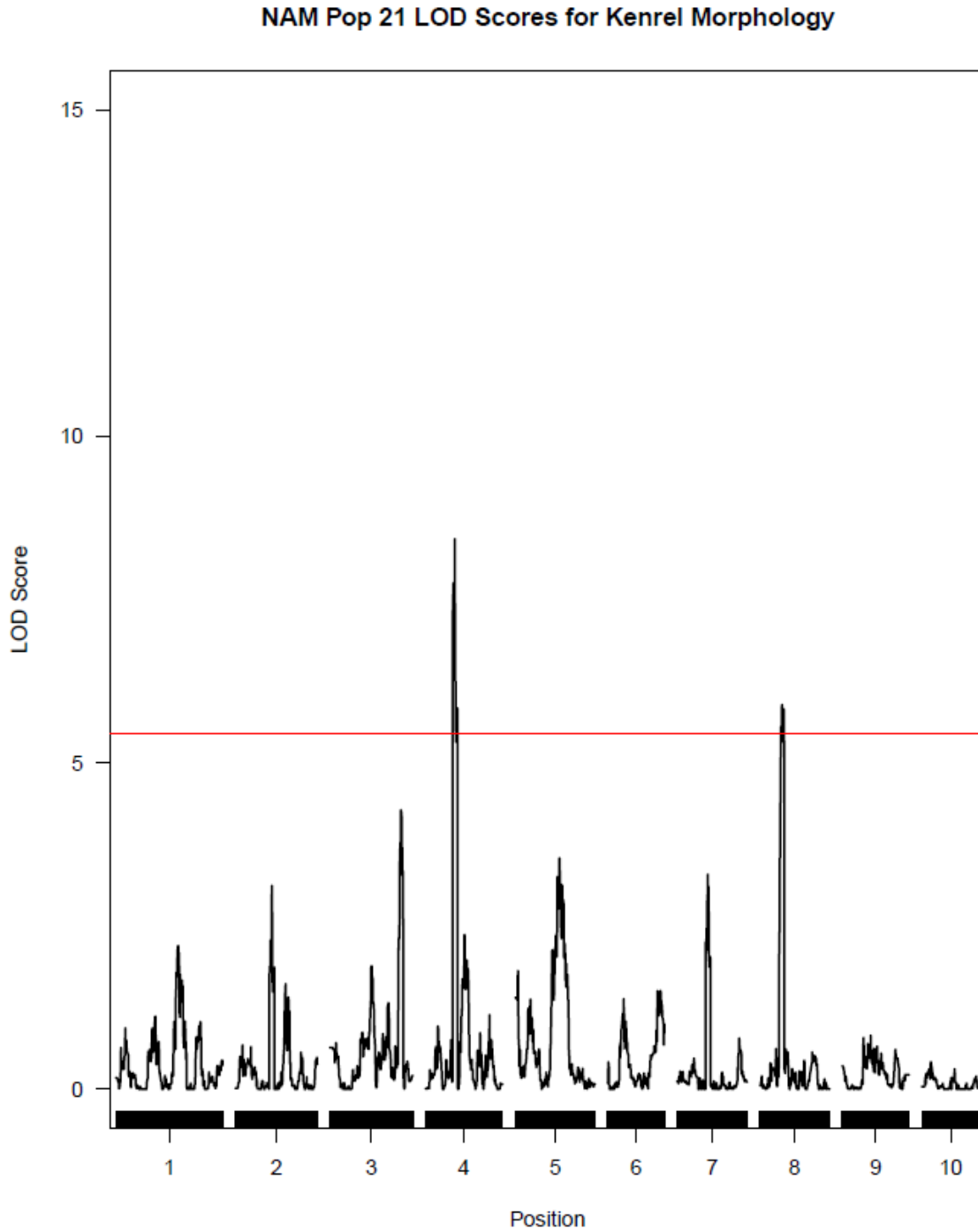

NAM Pop 21 LOD Scores for Kenrel Morphology

Figure S38. NAM population 22 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
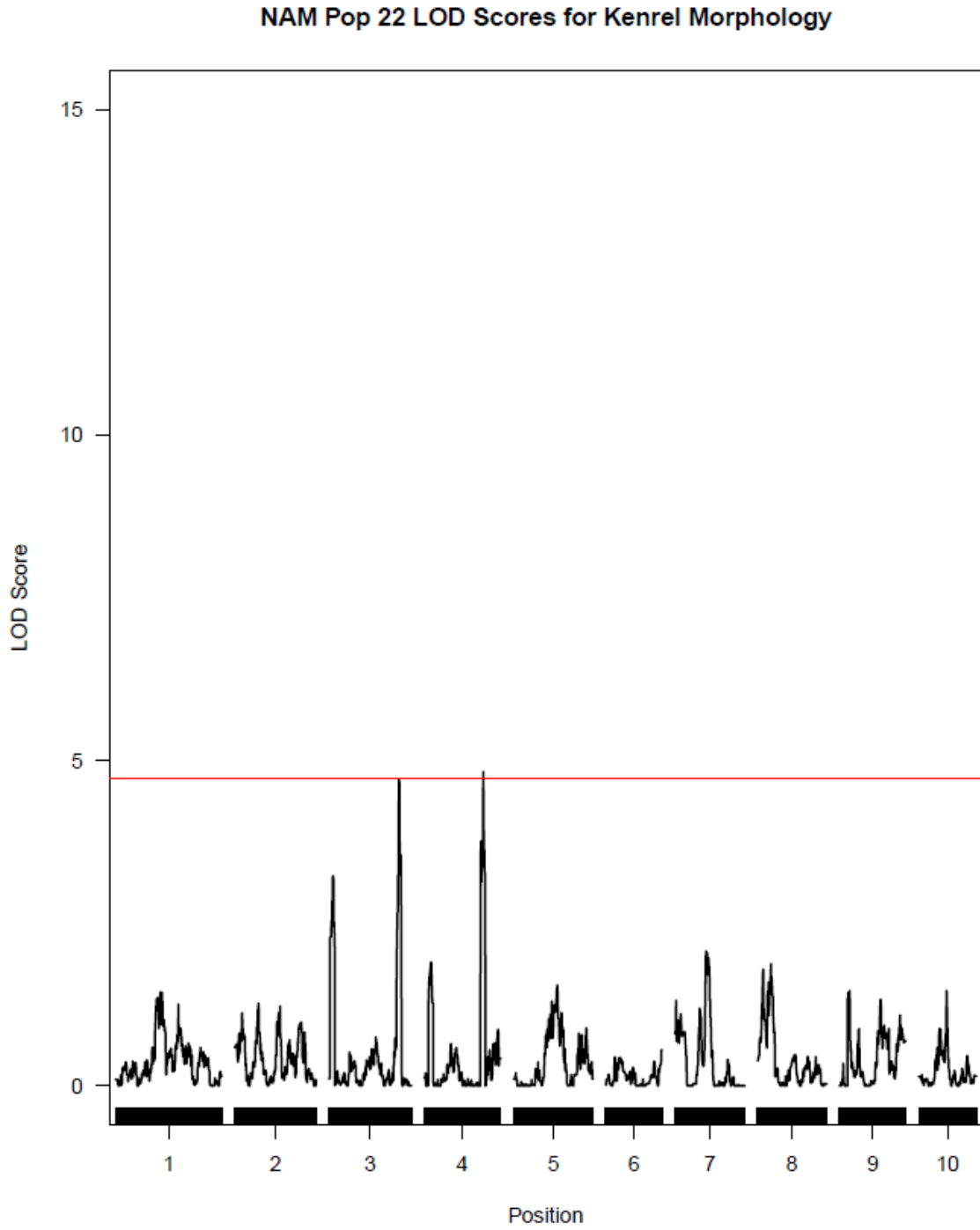

NAM Pop 22 LOD Scores for Kenrel Morphology

Figure S39. NAM population 23 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
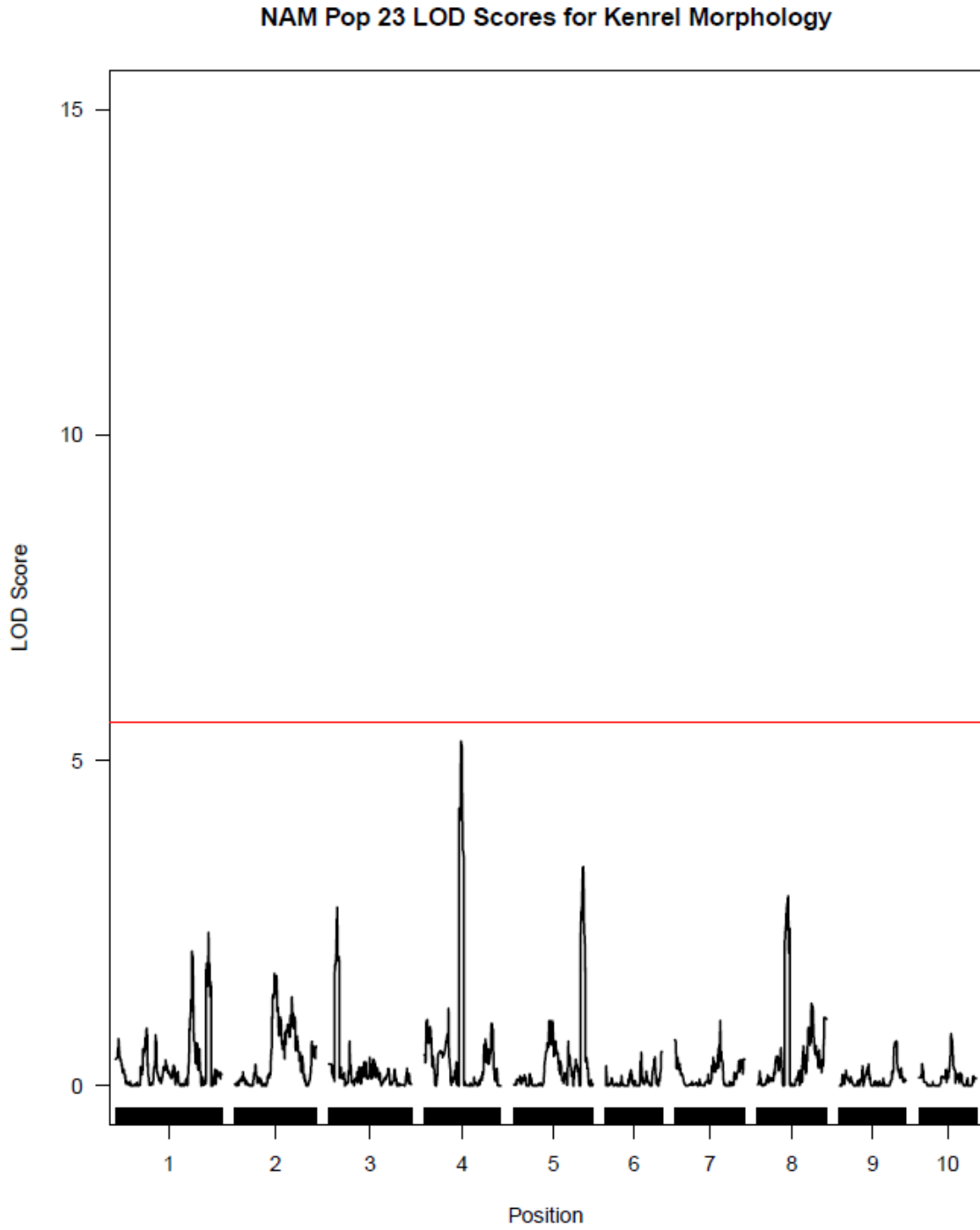
Figure S40. NAM population 25 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.
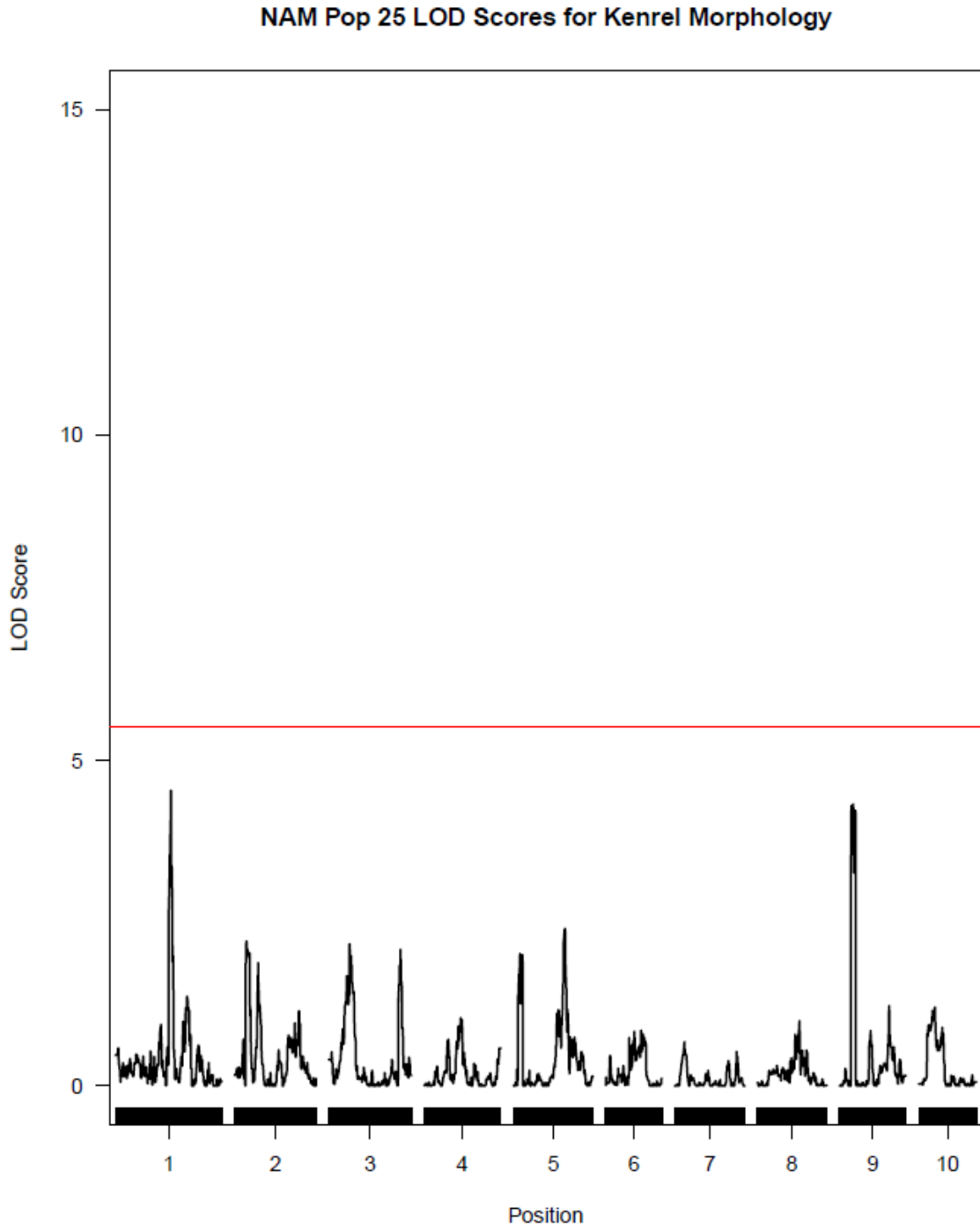

NAM Pop 25 LOD Scores for Kenrel Morphology

Figure S41. NAM population 26 linkage mapping for kernel morphology results. X-axis is position across the genome by chromosome. Y –axis is LOD score. Red line is significance threshold calculated by 100 permutations. Threshold values for each family are given in Table 1.

NAM Pop 26 LOD Scores for Kenrel Morphology