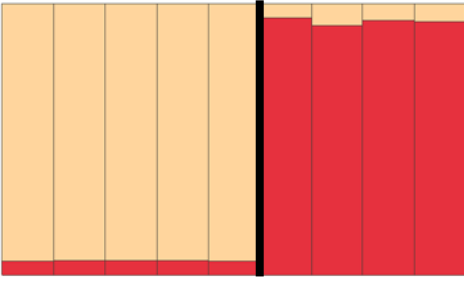


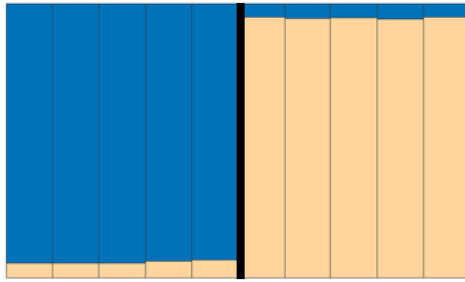
Supplementary Figure 1. Clustering results for individual breed. A) Heat-map style representation of 1127 structure runs at maximum values of K. Breeds are ordered as per Figure 1 on both the X and Y axes with thin colored lines representing each dog within the breed. Colors indicate the percentage of structure runs in which the two intersecting dogs clustered in the same groups. Red =>80% clustering, Orange 60-80%, Yellow 40-60%, Lt. Blue 20-40% and Dk. Blue <20%. The majority of dogs form unique clusters with all other dogs in the same breed as evidenced by the red blocks on the diagonal. Fourteen breeds do not form a solid red block on the diagonal. These breeds have larger than expected distances between individuals indicating a less homogeneous population structure (see Supplementary Table 3). Closely related breeds form red blocks off the diagonal. Related breeds are indicated by orange blocks off the diagonal. Thin black lines are placed every 12 breeds to aid in location of specific pairs.

B) Box-plot showing the distribution of distances within (1) and across (2) breeds.



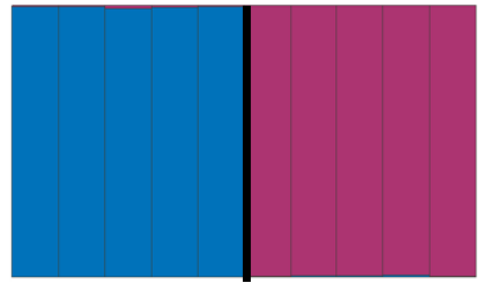
Boston Terrier

French Bulldog



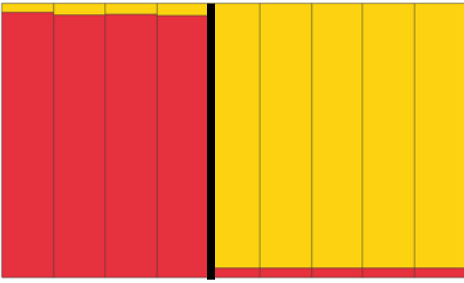
American Staffordshire Bull Terrier

Boston Terrier



American Staffordshire Bull Terrier

Kerry Blue Terrier



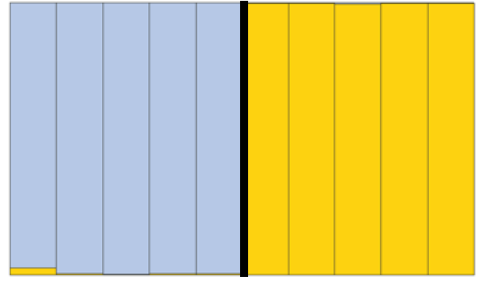
French Bulldog

Staffordshire Bull Terrier



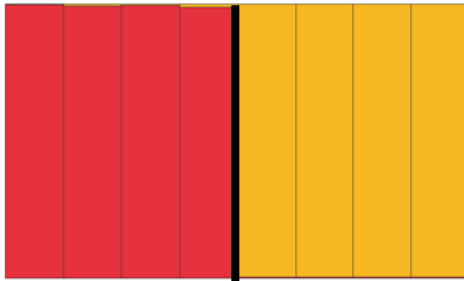
American Staffordshire Bull Terrier

Staffordshire Bull Terrier



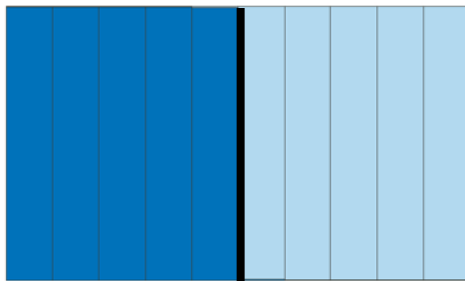
Miniature Bull Terrier

Staffordshire Bull Terrier



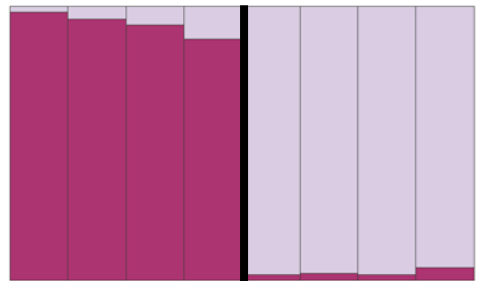
French Bulldog

Soft Coated Wheaten Terrier



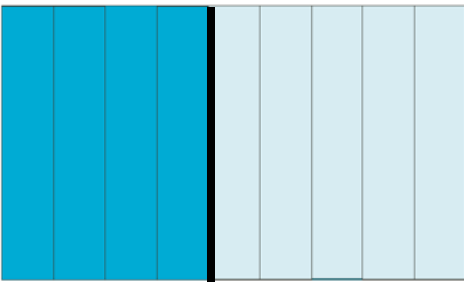
American Staffordshire Bull Terrier

Bulldog



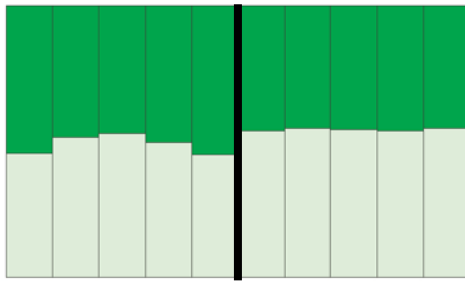
Manchester Terrier

Toy Manchester Terrier



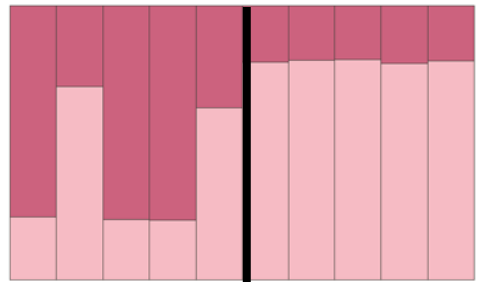
Scottish Deerhound

Irish Wolfhound



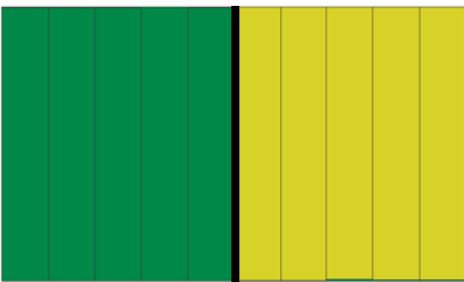
Grand Basset Griffon Vendeen

Petit Basset Griffon Vendeen



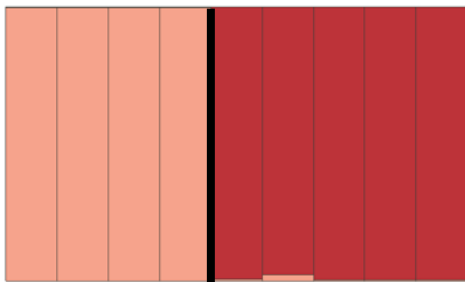
Miniature Poodle

Toy Poodle



Chow Chow

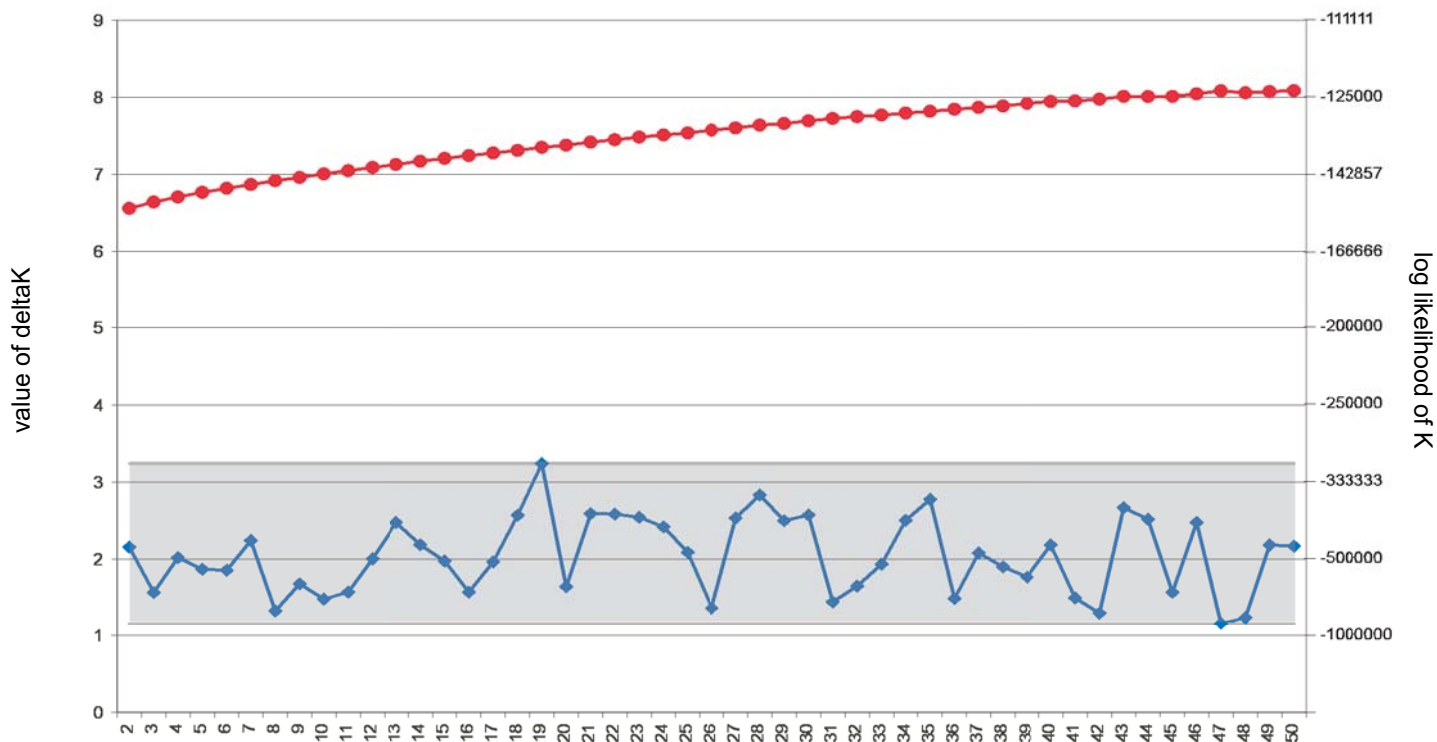
Shiba Inu



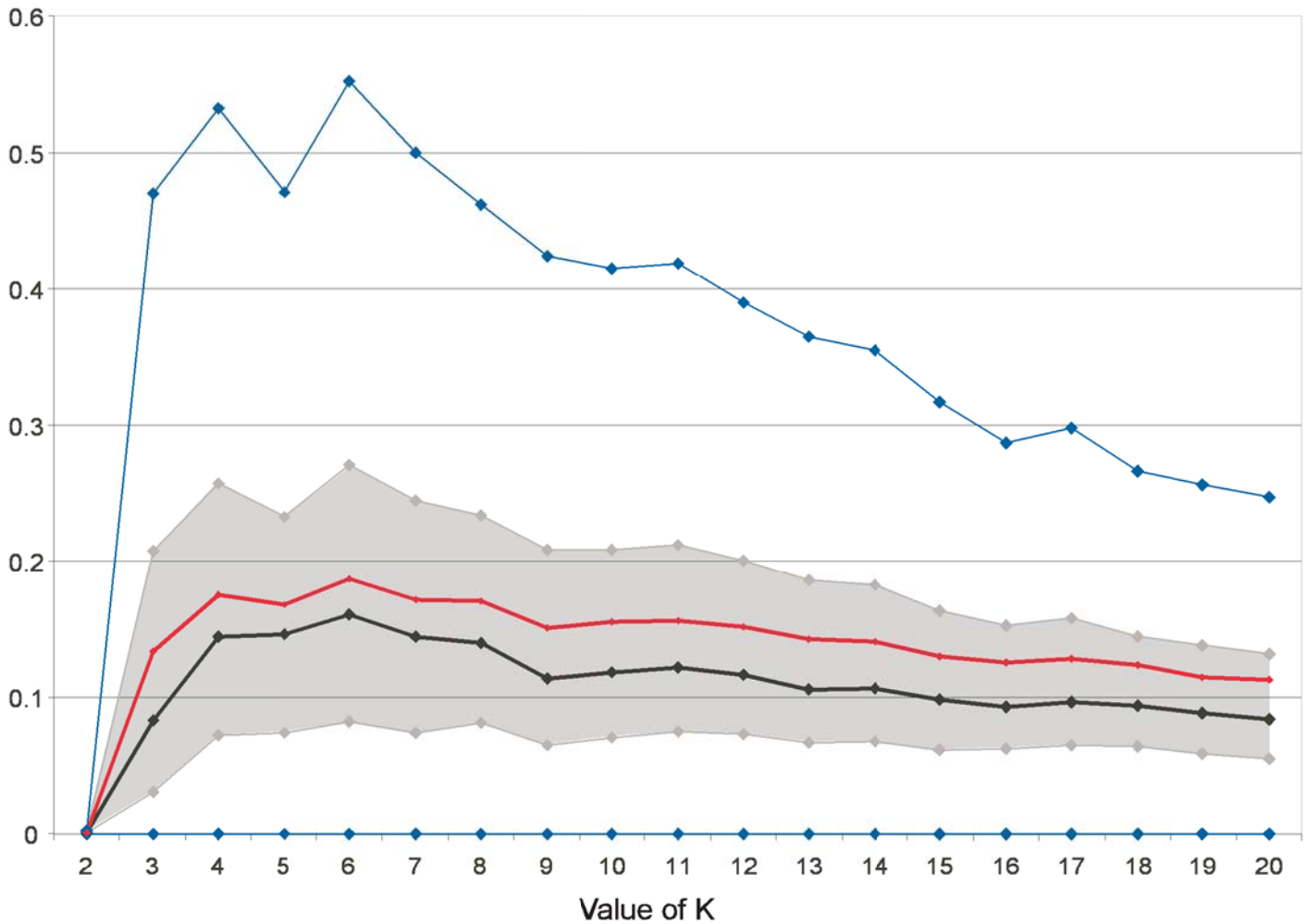
Norfolk Terrier

Norwich Terrier

Supplementary Figure 2. Clustering results for pairs of related breeds. Structure results averaged over 10 runs at $k=2$ for each pair of related breeds listed in Table 2 that have not been previously identified. Each dog is represented by a vertical bar divided into 2 colors representing the two possible populations. The breeds in each pair are separated by a thick black line. Breed names are listed below each graph.



Supplementary Figure 3. Graph of ΔK and log likelihood averaged over 10 runs at each value of K. Red line is the log likelihood calculation made by the program *structure*. Blue line is the value of ΔK . Shaded gray region shows the 95% confidence interval for ΔK . No points lie outside this interval. ΔK was calculated according to the methods of Evanno *et al* (2005).



Supplementary Figure 4. Standard deviations of clustering distances across 10 runs of structure for each value of K from 2 through 20. Euclidean distances are calculated for every pair of dogs based on their cluster assignment in each individual run. Standard deviation (SD) is then calculated from the variation in distances for each pair over 10 runs of structure. The Black line indicates the median SD for all possible pairs at a single value of K. The red line is the mean. The grey shaded area indicates the first and third quartile of the distribution of SDs. The dark blue lines indicated the upper and lower boundaries of the 95th % confidence interval of the distribution.

Supplementary Figure 5. A) Population structure of 638 dogs representing 132 breeds. Each graph is comprised of 638 vertical bars, each representing a single dog. The bar is divided into K colors, where K is the number of populations assumed. Two graphs are shown for K=5 representing two equally prevalent outcomes. B) Heat-map representation of structure results at K=2 through K=5 showing the gradual separation of the 5 groups. Breeds are ordered on the X and Y axes identical to the structure graphs above and in Figure 1.

Supplementary Table 1. Breeds that comprise the 638 dogs in the clustering analysis.

Breed	Number sampled	Clustering Integrity ¹
Afghan Hound	5	98.8
Airedale Terrier	4	99.3
Akita	5	99.6
Alaskan Malamute	5	95.7
American Cocker Spaniel	5	94.5
American Eskimo Dog	4	90.5
American Hairless Terrier	5	93
American Staffordshire Terrier	5	94.4
American Water Spaniel	5	80.5
Australian Cattle Dog	5	92.6
Australian Shepherd	5	83.3
Australian Terrier	5	93.7
Basenji	5	82.3
Basset Hound	4	98.9
Beagle	5	80.2
Bearded Collie ²	5	62.4
Bedlington Terrier	4	99
Belgian Malinois	5	88.7
Belgian Sheepdog	5	98.8
Belgian Tervuren	4	97
Bernese Mountain Dog	5	97.9
Bichon Frise	4	93.2
Bloodhound	5	98.5
Border Collie	5	93.1
Border Terrier	5	98.1
Borzoi	5	95.6
Boston Terrier	5	97.2
Bouvier des Flandres	4	95.7
Boxer	5	97.1
Brittany	5	93.9
Brussels Griffon	5	97.6
Bulldog	5	98.3
Bullmastiff	5	96.8
Cairn Terrier	5	92.1
Cardigan Welsh Corgi	5	96.8
Cavalier King Charles Spaniel	5	99.3
Cesky Terrier	5	97.2
Chesapeake Bay Retriever	5	91.7
Chihuahua	5	65.6
Chinese Crested	5	94.5
Chinese Shar-Pei	5	90.9
Chow Chow	5	97.5
Clumber Spaniel	5	99.2
Collie	5	99
Curly-Coated Retriever	4	88.5
Dachshund - standard, smooth	5	97.6

Dalmatian	4	96.9
Dandie Dinmont Terrier	5	97.1
Doberman Pinscher	5	98
English Cocker Spaniel	5	95.6
English Setter	5	95.9
Field Spaniel	5	97.9
Flat-coated Retriever	5	95.2
French Bulldog	4	97
German Shepherd Dog	5	98.6
German Short-haired Pointer	5	52.7
Giant Schnauzer	5	86.9
Glen of Imaal Terrier	5	97.1
Golden Retriever	5	88.6
Gordon Setter	5	88.7
Grand Basset Griffon Vendéen	5	95.2
Great Dane	5	96
Great Pyrenees	5	96.6
Greater Swiss Mountain Dog	5	98.6
Greyhound	5	94.7
Ibizan Hound	5	97.5
Irish Setter	5	97.2
Irish Terrier	4	89.2
Irish Water Spaniel	5	94.7
Irish Wolfhound	5	98.9
Italian Greyhound	5	97.4
Jack Russell Terrier	5	74.3
Keeshond	5	97.7
Kerry Blue Terrier	5	95.6
Komondor	5	95.1
Kuvasz	5	95.6
Labrador Retriever	5	89.4
Lancashire Heeler	5	94
Leonberger	4	98.2
Lhasa Apso	5	89.8
Maltese	5	89.5
Manchester Terrier - Standard	4	98.8
Manchester Terrier - Toy	4	95.1
Mastiff	5	99.1
Miniature Bull Terrier	5	99
Miniature Pinscher	5	78.2
Miniature Poodle	5	94.7
Miniature Schnauzer	5	96.6
Newfoundland	5	89.1
Norfolk Terrier	4	98.4
Norwegian Elkhound	5	91.2
Norwich Terrier	5	99.3
Nova Scotia Duck Tolling Retriever	5	98.1
Old English Sheepdog	5	94.7
Papillon	5	82.2
Pekingese	5	97.5
Pembroke Welsh Corgi	5	96.7

Petit Basset Griffon Vendeen	5	92.6
Pharaoh Hound	4	98.2
Pointer	5	98.1
Pomeranian	5	88.3
Portuguese Water Dog	5	80.5
Presa Canario	5	66.4
Pug	5	99.3
Puli	4	89.3
Rhodesian Ridgeback	5	87.1
Rottweiler	5	97.8
Saint Bernard	5	94
Saluki	5	82.2
Samoyed	5	98.2
Schipperke	4	98.1
Scottish Deerhound	4	99
Scottish Terrier	5	97.7
Shetland Sheepdog	5	98.2
Shiba Inu	5	94.7
Shih Tzu	5	96.1
Siberian Husky	5	90.3
Silky Terrier	5	96.7
Soft Coated Wheaten Terrier	4	99
Spinone Italiano	5	89.2
Staffordshire Bull Terrier	5	96
Standard Poodle	5	95.4
Standard Schnauzer	5	97.3
Tibetan Terrier	5	91.4
Toy Poodle	5	94.8
Vizsla	4	94.7
Weimaraner	5	93.7
Welsh Springer Spaniel	5	98.3
Welsh Terrier	4	66.9
West Highland White Terrier	5	95.7
Whippet	5	96.7
Wirehaired Pointing Griffon	5	92.1

1) Clustering integrity is the percentage of 1127 runs in which all members of the breed cluster identically. The four breeds with the lowest within breed clustering have members that cluster consistently with other breeds more than 80% of the time. These include: four German Shorthair Pointers, four Welsh Terriers, one Presa Canario, and one Jack Russell Terrier.

2) Four of five Bearded Collies cluster in 97.1% of all runs. The fifth always clusters away from these four and is suspected to be a mis-labeled Polish Lowland Sheepdog.

Supplementary Table 2. Nine new microsatellite markers located within the previously identified Collie Eye Anomaly zero recombination interval (Lowe *et al.*, 2003), flanked by genes AAMP and EP4A4 on CFA37. The CanFam2 start position identifies the 5-prime end of the PCR product. These microsatellites, derived by interrogating the TIGR 1.5 x canine sequence database using human sequence for genes located between AAMP and EP4A4 on HSA2, and subsequently located in the 7.5x assembled CanFam2 sequence, form several variant haplotypes among *cea*-affected Shetland Sheepdogs (SS), Collies, Australian Shepherds (AS) and Border Collies (BC). The variation in these haplotypes is compatible with derivation from a common ancestor by recombination events in either the telomeric or centromeric end.

MS#	Name	Forward Primer	Reverse Primer	Start Position CanFam2	Repeat Type	Alleles in affected dogs															
						SS	Collie 1	Collie 2	Collie3	AS	BC 1	BC 2	BC 3								
1	FH4617	ACACATCAAATGAACCTTTATCA	GGGAATGGAAAAAGTAAGCTAAA	28,172,980	CA	1	2	2	2	2	2	2	2	3	3	2	2	3	2	2	2
2	FH4471	TGCGTCATGAAACGGTATCT	GGTGGCTCAGTCGGTTAAGT	28,345,309	CT	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	FH4619	CTCAAGAATCACATCAAGACCTC	GAAGTGCAGGAGTTAGGTAGTCC	28,458,370	TAAA CAAA	5	5	5	5	5	5	4	5	5	5	5	5	5	5	4	5
4	FH4620	TTCTCTGGGAATAGGTGAAGG	TAATCTGGCTCAAGGGAAAGT	28,651,711	CA	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2
5	FH4621	CCTCCAAAATACACTCCATT	GCTATCTGCTTACCACCTTCC	28,768,900	CA	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2
6	FH4622	TAAAGGCAGATAGGAGTGCTAGAT	AAGTCTATGATTTTAGGGTTGTGG	28,864,966	TGTA	2	2	2	2	2	2	2	2	2	3	1	1	1	1	1	1
7	FH4306	GTGGTAGGGTGCTCAAGGAA	AATCCACAGACCGTAAAGG	29,034,333	GAAA	3	3	3	3	3	3	3	3	5	4	5	3	4	1	3	3
8	FH4625	GTTAAAGGCTCTGTGACCACTC	ATAGTTGGCTTGCTTGTTTGT	29,326,364	TAAA CAAA	3	3	3	3	3	4	3	3	4	4	3	3	3	4	1	1
9	FH4615	TGGTGTAGGGTTTGGTGACA	TCTAGCAATGTGGCAAGGTG	29,354,908	GAAA	1	2	1	2	1	1	1	2	2	2	1	1	1	3	1	1

Supplementary Table 4. Breeds with known incidence of colobomas¹.

Breeds and varieties segregating molecularly defined CEA (n = 9)

Australian Shepherd, Border Collies, Boykin Spaniel, Lancashire Heeler (Ormskirk Terrier), Longhaired Whippet/ Silken Windhound, Nova Scotia Duck Tolling Retriever, Rough Collie, Shetland Sheepdog, Smooth Collie

Breeds with both Choroidal Hypoplasia and Colobomas, potentially but not yet proven to be CEA (n-18)

Bearded Collie, Cairn Terrier, Cavalier King Charles Spaniel, Clumber Spaniel, Cocker Spaniel (American), Curly-Coated Retriever, Doberman Pinscher, German Shorthaired Pointer, Golden Retriever, Havanese, Lhasa Apso, Lowchen, Miniature Schnauzer, Old English Sheepdog, Poodle (Standard, Miniature, Toy), Samoyed, Siberian Husky, Tibetan Terrier

Breeds with Staphyloma/Coloboma reported, but likely to be part of Merle phenotype (n = >3)

Catahoula Leopard Dog, Dachshund (All Varieties), Great Dane

Breeds with sporadic isolated colobomas (n=77)

Afghan Hound, Airedale Terrier, Akita, Alaskan Malamute, American Eskimo, American Staffordshire Terrier, Australian Cattle Dog, Basenji, Beagle, Bedlington Terrier, Belgian Sheepdog, Belgian Tervuren, Bernese Mountain Dog, Bichon Frise, Bloodhound, Borzoi, Bouvier Des Flandres, Boxer, Boykin Spaniel, Briard, Brittany Spaniel, Brussels Griffon, Bullmastiff, Cardigan Welsh Corgi, Chesapeake Bay Retriever, Chihuahua, Chinese Crested, Coton De Tulear, English Cocker Spaniel, English Springer Spaniel, English Toy Spaniel, Flat-Coated Retriever, German Shepherd, Glen Of Imaal Terrier, Gordon Setter, Great Pyrenees, Harrier, Ibizan Hound, Irish Setter, Irish Wolfhound, Italian Greyhound, Jack Russell Terrier, Japanese Chin, Keeshond, Labrador Retriever, Leonberger, Mastiff (English), Miniature Australian Shepherd, Miniature Bull Terrier, Miniature Pinscher, Newfoundland, Norfolk Terrier, Norwich Terrier, Papillon, Pembroke Welsh Corgi, Petit Basset Griffon Vendeen, Pomeranian, Portuguese Water Dog, Pug, Rhodesian Ridgeback, Rottweiler, Saluki, Sealyham Terrier, Shih Tzu, Soft Coated Wheaten Terrier, Sussex Spaniel, Swedish Vallhund, Tibetan Spaniel, Welsh Springer Spaniel, West Highland White Terrier.

1) Data extracted from: "Ocular Disorders Proven or Suspected to be Inherited in Purebred dogs", American College of Veterinary Ophthalmologists, 2007

Supplementary Table 5. Canine BAC clones comprising a contig covering the *cea* interval.

A. Primers used for generation of gene-specific probes for screening the canine RPC181 BAC library, to identify BAC clones that map to the *cea* interval on CFA37.

GENE	Forward primer	Reverse primer
<i>CRYBA2</i>	tcagaagatcagagaattccttcc	cagtccgaactttccgagattcgt
<i>CDK5R2</i>	aggtcttgaaagacctgCGTgaaga	tgctgCGTgCCTgggCGacttcgt
<i>NHEJ1</i>	ggttgGagaatcctttctctg	gaaggTactaactgtacaacc
<i>SLC23A3</i>	atccaaactctgacctcttacac	caggatggtggacacaccacag
<i>ABCB6</i>	gtctcacattggagttgtgc	tctcgTtagacgtatccagc

B. Canine BAC clones; the gene specific probes used to identify them, and the SNP amplicons amplified within each clone.

BAC Clone identification	Gene probe(s) for which clone is positive	CanFam2 location of end sequences*	SNPs
96-N05	<i>CRYBA2</i>	28.379 - 28.558	rs23979435, rs24023522, rs24023523, rs23998244, rs24023545
59-K05	<i>CDK5R2</i>	28.405 - 28.559	rs23979435, rs24023522, rs24023523, rs23998244, rs24023545
318-F08	<i>NHEJ1</i>	28.527 - 28.713	rs24009189, rs23961769, rs8535216, rs24031946, rs8535215, rs8797935, rs23993534
96-B02	<i>NHEJ1, SLC23A3</i>	28.616 - 28.774	rs24009189, rs23961769, rs8535216, rs24031946, rs8535215, rs8797935, rs23993534, rs23961971
262-E17	<i>NHEJ1</i>	28.697 - 28.838	rs23979435, rs24023523, rs23998244, rs24023545, rs23961971
222-K13	<i>ABCB6</i>	28.754 - (sequence failed)	None

* Mb position on CFA37 in CanFam2 assembled canine genome sequence

Supplementary Table 6. Locations and sequence information for the conserved transcription factor binding domains identified within the cea-associated deletion (see Figure 4.)

<u>Transcription factor name</u>	<u>Transcription factor class</u>	<u>Sequence</u>	<u>position¹</u>	<u>Score</u>	<u>Strand</u>
<i>TBP</i>	TATA-binding protein	GCGGGCTCTTTATTT	32-46	7.816	-
<i>HFH-3</i>	Forkhead	GGCTCTTTATTT	35-46	10.000	+
<i>TBP</i>	TATA box binding protein	GCTCTTTATTTATTT	36-50	6.555	-
<i>HFH-1</i>	Forkhead	CTTTATTTATT	39-49	7.914	+
<i>HFH-2</i>	Forkhead	CTTTATTTATTT	39-50	11.054	+
<i>HFH-3</i>	Forkhead	CTTTATTTATTT	39-50	10.811	+
<i>HNF-3beta</i>	Forkhead	CTTTATTTATTT	39-50	11.951	+
<i>SQUA</i>	MADS-box	CTTTATTTATTTAT	39-52	8.671	-
<i>Broad-complex_4</i>	ZN-FINGER, C2H2	TTTATTTATTT	40-50	9.487	-
<i>FREAC-2</i>	Forkhead	TTTATTTATTTATC	40-53	9.012	-
<i>Hunchback</i>	ZN-FINGER, C2H2	TTATTTATTT	41-50	6.384	-
<i>Athb-1</i>	HOMEO-ZIP	TATTTATT	42-49	7.226	+

1) Position given is in reference to the alignment in Figure 4.