The Game Changers

Our sport's legacy of inheritance as documented by all of the world's libraries of revered stud records is being challenged by an awesome new world of technology.

by Rommy Faversham

Perhaps it's time for someone to just come out and say it. Modern thoroughbred pedigrees, more commonly than not, have become 'less interesting' over time as runners across the globe continue to become more and more alike.

Don't blame the female families. The good ones are plentiful and as diverse as ever. But what of the constant decimation suffered by the once broad spectrum of sire lines? Fashion and search for profit has always been the norm. In recent times however, there has been a profound (almost logarithmic) reduction in genetic diversity, particularly in regards to the Y chromosome and its unique contribution to the animal's total genome.

In the meantime, genetics labs around the world are already well into the identification of segments within a runner's DNA most relevant to their racing performance. Some of these studies have also uncovered a significant number of important thoroughbred pedigrees from the past which feature striking inaccuracies. Issues of descent have, in fact, been identified in a majority of the most successful of female families.

Four years ago, a large group of veterinary doctors centered in Vienna reported on their Y chromosomal haplotyping to address the ancestry of key sire lines. To the astonishment of more than a few pundits, the researchers found that St. Simon, one of the nineteenth century's greatest stallions carried the oB1 Y chromosome indicating his tail-male descendancy was through foundational sire Herod, not Eclipse - as stud books had always guaranteed.

Another group had earlier studied the mitochondrial (D-loop) DNA extracted from the preserved jaw of Bend Or, supposed winner of the 1880 English Derby winner as well as tail-male ancestor to an overwhelming proportion of contemporary thoroughbreds. As it turns out, the subject's mDNA was shown to share a common sequence, designated L4a, which indicated he was indeed a descendant of the Martha Lynn family (#2-h) and therefore not Bend Or (out of Rouge Rose from family #1-k) but rather a different son of Doncaster named Tadcaster (ex Clemence).

Investigation into the thoroughbred's mitochondrial DNA is now well over ten years old with clearly defined family groups or haplogroups and their sub-groups or haplotypes. Renown pedigree authority and Bluebloods contributor Alan Porter has, in recent years, been referring to these distinct haplogroups and how often they repeat within the ancestry of important individuals. This is a very useful update to the powers of deep linebreeding to influential clans of the past.

Already, the game-changing (no, life-changing) development of artificial intelligence equipped with comprehensive analysis of the entire thoroughbred genome figures to catapult the grand art of pedigree analysis into something extremely different. The same number of trophies will be awarded but they'll be sought after with whole new game plans. Nicks and inbreeding patterns (oh, no) will be replaced by dossiers filled with information regarding the presence or absence of specific gene sequences known to effect racing ability in some applicable manner.

The Speed Gene

High Tech's first major inroad into the breeding of thoroughbreds was announced by the media in 2012 following publication in the scientific journal Nature Communications of work provided by an international collection of researchers led by UC Dublin professor Emmeline Hill and Dr. Mim Bower, an archaeo-geneticist at the University of Cambridge. This team combined to identify and then fully describe what has quite appropriately come to be known as the speed gene.

The Myostatin (MSTN) gene group normally acts to limit skeletal muscle mass by regulating both the number and growth of muscle fibers. Using

TABLE 1 - Historic Thoroughbred specimen details from Hill, Bower, et al. (2012)

Foundation ancestor	Sire	Element anlayzed	MSTN genotype
BEND OR (1877)	Doncaster	mandible	T / T
CORRIE ROY (mare; 1878)	Galopin	metacarpus	T / T
DONOVAN (1886)	Galopin	humerus	T / T
ECLIPSE (1764)	Marske	humerus	T / T
HERMIT (1864)	Newminster	tooth	T / T
HYPERION (1930)	Gainsboroough	tooth	T / T
ORMONDE (1883)	Bend Or	metacarpus	T / T
PERSIMMON (1893)	St. Simon	mandible	T / T
POLYMELUS (1902)	Cyllene	tooth	T / T
ST. FRUSQUIN (1893)	St. Simon	metacarpus	T / T
ST. SIMON (1881)	Galopin	mandible	T / T
STOCKWELL (1849)	The Baron	metacarpus	T / T
WILLIAM THE THIRD (1898)	St.Simon	metacarpus	T / T

a combination of molecular and pedigree-based approaches, the scientists traced the haplogroup backwards in time, studying the MSTN segments of hundreds of runners, including the world's racing elite. DNA from other breeds of horses as well as distinct pieces from the skeletons of twelve historic progenitors were also evaluated (Table 1).

The key finding was the presence of variation at the MSTN locus. This included not only the ancestral 'wild-type' variant or T-allele but also a second cardinal variant, the result of a mutation, the C-allele which contributes to morphological changes associated with muscle hypertrophy shown to be particularly pronounced in two year-olds in training.

This new bivalent genetic marker is then tested as to the likelihood of early maturity as well as greater height and muscle mass in thoroughbreds carrying the C/C allele combination of the MSTN gene. Individuals with copies of both alleles, C/T (heterozygous) figure more in the middle of the spectrum as to size, time of maturity, and optimal racing distance, while those with the T/T genotype tend to be smaller, later-maturing types with more ability over a longer course of ground.

In Table 2, the frequencies of the C and T alleles and their genotypes demonstrated a high level of concordance with the proven ability to sprint vs. route. Among 69 individuals who won a Group and/or Listed stakes in England, Ireland, New Zealand and/or the United States at a distance of seven furlongs or less, the ratio of C allele to T allele was seven to three (0.70/0.30). This is almost the reversal (0.34/0.66) of the C/T ratio calculated from 96 elite runners from the aforementioned countries who captured at least one Group and/or Listed stakes at nine furlongs or beyond.

Another interesting finding noted in Table 2 is the significantly higher proportion of speed genes in Australia. The data suggests an overall C/T ratio that is almost at equivalence (0.51/0.49) for many of the world's major racing centers. Down Under, however, it appears to be closer to 0.64/0.36. This ratio of close to two to one may reflect an ongoing emphasis towards precocity and speed domestically.

Genotypes of the proven stakes-winning sprinters were 46% C/C, 46% C/T with just 7% T/T (homozygous for the wild-type T-allele). Conversely, proven stakes-winning routers' genotypes for the MSTN gene was 35% T/T, 61% C/T with just 3% C/C. Later studies replicated and validated this original research on a massively larger scale. One subsequent study found 83% of European runners with the C/C genotype had an optimum race distance of a mile or shorter while more than 89% with the T/T genotype from the sane population ran best over distances greater than a mile.

Table 2 also shows the speed gene was not identified in non-horse equids, such as donkeys or zebras. It also was not apparent in Standardbreds yet dominated MSTN genotypes in Quarter Horses by an overwhelming nine to one ratio. The horse breed carrying the highest frequency of the C variant in this study was shown to be Shetland, one of the smaller, British types who appears to have made some key contributions to the thoroughbred genome.

It is the result of human selection that this new and extremely rare gene variant ultimately became common. In the late seventeenth and early eighteenth centuries, during the early years of thoroughbred development, breeders were focused on stamina for the longer distances of the day and thus, quite inadvertently, maintained very high levels of the T allele within this new elite population. This was confirmed by the researchers when all thirteen of the tested historic runners and progenitors were shown to carry the T/T MSTN genotype (Table 1).

Conditions began to change by the second half of the nineteenth century as more races were run at shorter distances with a greater number of runners starting their careers as two year-olds. As a result, the new selection process favors individuals with C/C or C/T genotypes who are much more likely to develop the musculature necessary to sprint with intense high-speed bursts. It may indeed be the T-allele that now needs replenishment.

Further investigation provided some remarkable information regarding two of the key points in the paths of transmission taken by the speed gene. Firstly, the studies ascertained the C-allele was introduced during the new breed's foundational stages and, quite notably, as a one-time event. According to Dr. Bower, "our findings point to a British-native mare foaled around 300 years ago, who crossed with the three foundation stallions (Eclipse, Herod and Matchem) or their early [antecedents], as being the

TABLE 2 - Allele and Genotype frequencies for the Myostatin (speed) gene

	Sample size	Genotype frequency			Allele frequency	
		C/C	C/T	T/T	C	T
Equus asinus (donkey)	40	0.00	0.00	1.00	0.00	1.00
Equus quagga (zebra)	2	0.00	0.00	1.00	0.00	1.00
Shetland (Scotland)	16	0.38	0.25	0.38	0.50	0.50
Shetland (Sweden)	42	0.29	0.40	0.31	0.49	0.51
Quarter Horse	35	0.83	0.14	0.03	0.90	0.10
Standardbred	63	0.00	0.00	1.00	0.00	1.0
Contemporary SWs in GB/IRE/NZ/USA at 1,400 m. or less	69	0.46	0.46	0.07	0.70	0.30
Contemporary SWs in GB/IRE/NZ/USA at 1,800 m. or more	96	0.03	0.61	0.35	0.34	0.66
Contemporary SWs in GB/IRE/NZ/USA at all distances	207	0.22	0.57	0.21	0.51	0.49
Contemporary SWs in Australia at all distances	123	0.38	0.51	0.11	0.64	0.36

* from Nature Communications, January 24, 2012, Hill, Bower, et al

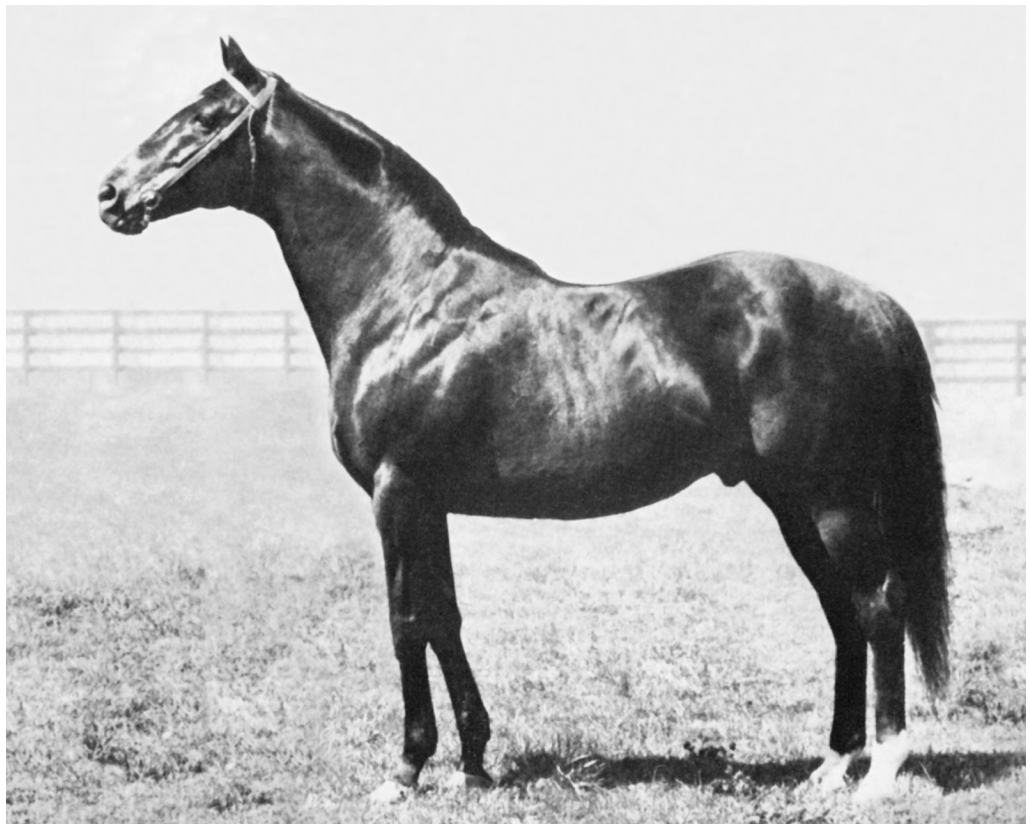
most likely source of the original C-gene for speed."

For much of the next two hundred and fifty years plus, the speed gene maintained a relatively very low frequency level but figures to have thrived in its limited opportunities at shorter distances. Otherwise, the speed gene's specific paths of transmission from genesis to modern times remains unknown.

The study was able to establish that the producer most responsible for the dramatic proliferation of the C-allele over the past sixty years or so was the Nearco stallion Nearctic (1954), known of course for siring the great Northern Dancer. The data would show contemporary racehorses with the C/C genotype traced back to Nearctic in significantly greater measure than Nearco's two other breed-shaping sire sons Nasrullah and his three-quarter brother Royal Charger, both undisputed Brilliant chefs-de-race.

At this point I cannot help but digress, for in the 1990s, the classification of chefs-de-race (breed-shaping stallions) had somehow become a hot-button issue and fights could break out among pedigree mavens over some pretty esoteric stuff. One of the noisiest of these was played out on the pages of America's Daily Racing Form and Blood-Horse over much of 1993 and '94 after yours truly, as member of a Dosage advisory panel, wrote a paper recommending the classification of Nearctic to Brilliant chef-de-race status. The subsequent unanimous 7-0 vote meant most pedigrees would have slightly higher dosage figures reflecting an additional measure of speed to be accounted for. What I failed to take into account is that the move could disrupt well-known dosage pioneer Dr. Steve Roman's perfect record of Kentucky Derby winners with an index of 4.00 or under. Having long spurned joining our panel, Roman immediately went on the offense. "This work is cause for alarm" he wrote in the DRF. "It is an outrageous perversion of the principles and standards of dosage as understood by the racing public". And with that spirited flurry, it seemed as if everyone with a turf column chimed in. The Nature Communications article was a nice surprise almost twenty years later. Another example of what DNA evidence can provide years later.

Nearctic's five generation pedigree features three strains of Canterbury Pilgrim highlighted in bold red. As such, he is himself an RF. Important sires within his pedigree proven to carry the T/T genotype for the MSTN gene



Nearctic by Nearco won 21 races frpm 47 starts mostly at sprint distances and his male line is now the dominant one for speed in the modern thoroughbred.

represent pathways that cannot convey the C-allele. These include Nearctic's paternal great-great grandsire Polymelus, the three strains of St. Simon that appear in his pedigree's fifth generation, his broodmare sire Hyperion and Hyperion's broodmare sire St. Frusquin (all highlighted in bold blue). From where then did Nearctic inherit his speed gene? The answer figures to require more MSTN genotyping of important past sires and dams whose skeletal specimens are still archived. Hopefully somewhere, there is a sample from beyond a century ago of a key thoroughbred who carried the C/T signature.

Last year, Darley stallion Cracksman was novelly advertised as C/C for the speed gene. World class from a mile and a quarter to a mile and a half, he had plenty of sources of speed in his pedigree to help explain the seemingly

NEARCTIC (CAN) Brown colt, 1954

NEARCO Br 1935	Pharos	Phalaris	**Polymelus**	Cyllene Maid Marian	9e 3f
			Bromus	Sainfoin Cheery	2g 1i
		Scapa Flow	1 **Chaucer**	**St Simon** **Canterbury Pilgrim**	11c 1g
			Anchora	Love Wisely Eryholme	11e 13e
	Nogara	Havresac	Rabelais	**St Simon** Satirical	11c 14a
			Hors Concours	Ajax Simona	2i 8h
		Catnip	Spearmint	Carbine Maid of the Mint	2h 1c
			Sibola	The Sailor Prince (GB) Saluda	19 4r
LADY ANGELA Ch 1944	**Hyperion**	Gainsborough	Bayardo	Bay Ronald Galicia	3o 10a
			Rosedrop	**St Frusquin** Rosaline	22b 2n
		Selene	1 **Chaucer**	**St Simon** **Canterbury Pilgrim**	11c 1g
			Serenissima	Minoru Gondolette	5c 6e
	Sister Sarah	Abbots Trace	Tracery	Rock Sand Topiary	4n 19a
			Abbots Anne	Right-Away Sister Lumley	11e 4j
		Sarita	1 Swynford	John o' Gaunt **Canterbury Pilgrim**	3e 1g
			Molly Desmond	Desmond Pretty Polly	16c 14c

incongruous genotype. The best of Cracksman's first crop, this season's unbeaten French Derby winner Ace Impact is stoutly bred so it will be interesting to see what was the intent as well as the results.

More important stuff on chromosomes

Another international team of geneticists also led by Dublin professor Emmeline Hill released an update of their work at the end of 2022 in which they identified seven genes, G6PC2, HDAC9, KTN1, MYLK2, NTM, SLC16A1 and SYNDIG1, which are shown to have central roles in muscle, metabolism, and neurobiology. Variation of these genes can be used to exploit genetic improvements within race horse populations who are in preparation for specific types of racing.

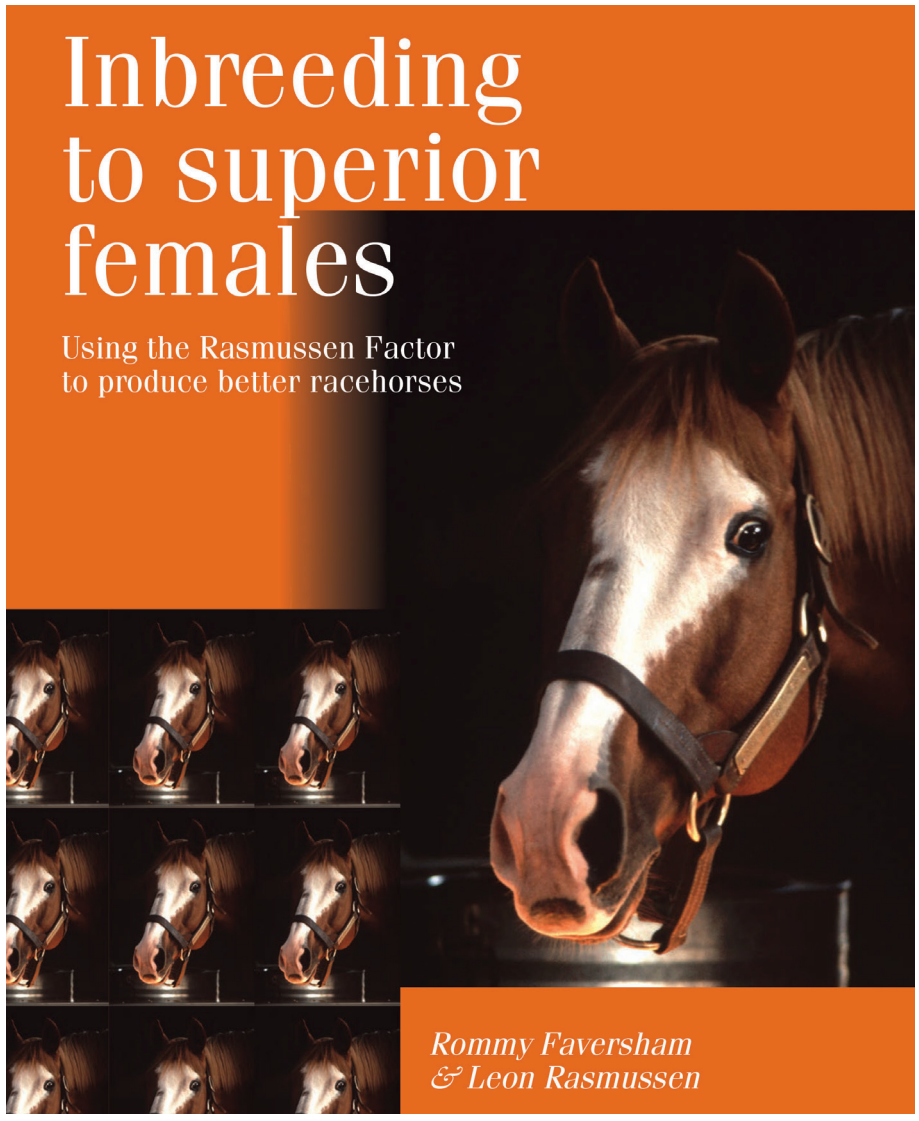
The G6PC2 gene encodes for a key enzyme in glucose metabolism. HDAC9 gene variants are associated with the maximal oxygen uptake (VO2max) in response to training. MYLK2 has a critical role in muscle contraction. NTM encodes for neurotrimin, which functions in brain development to regulate neural growth and synapse formation which, in turn, influences learning and memory. The NTM locus was found to be most closely associated with the number of racecourse starts. The protein produced by the SYNDIG1 gene regulates the development of excitatory synapses. These, then, are the new key drivers of the racing phenotype and they bring a whole new meaning to the expression, "checking all the boxes".

What about A.I.?

Artificial intelligence is very different from 'classic' laboratory science. It is the transfer of our abilities to perceive, synthesize and infer information to non-living operators, a promethean venture if there ever was one. Things get rather moot when the potential of A.I. is considered. Breeding, racing, handicapping - they will all certainly become something quite different.

For now, one can get a taste of artificial intelligence by downloading the app ChatGPT, already the fastest growing consumer software in history despite its well known tendency to 'confidently provide inaccurate information' (Wikipedia). When I decided to question it on some breeding theory the program arrived at the conclusion that, "the intentional practice of female family inbreeding is not considered a recommended approach". What do you know? Another critic.

HYPERION by Gainsborough. DNA analysis of his tooth revealed him to have the T/T MSTN genotype.



"Inbreeding To Superior Females" by Rommy Faversham (author of this article) and Leon Rasmussen is one of the best selling thoroughbred reference works ever published. Copies are still available from Bluebloods' office during business hours or on the website bluebloods.com.auin the modern thoroughbred.