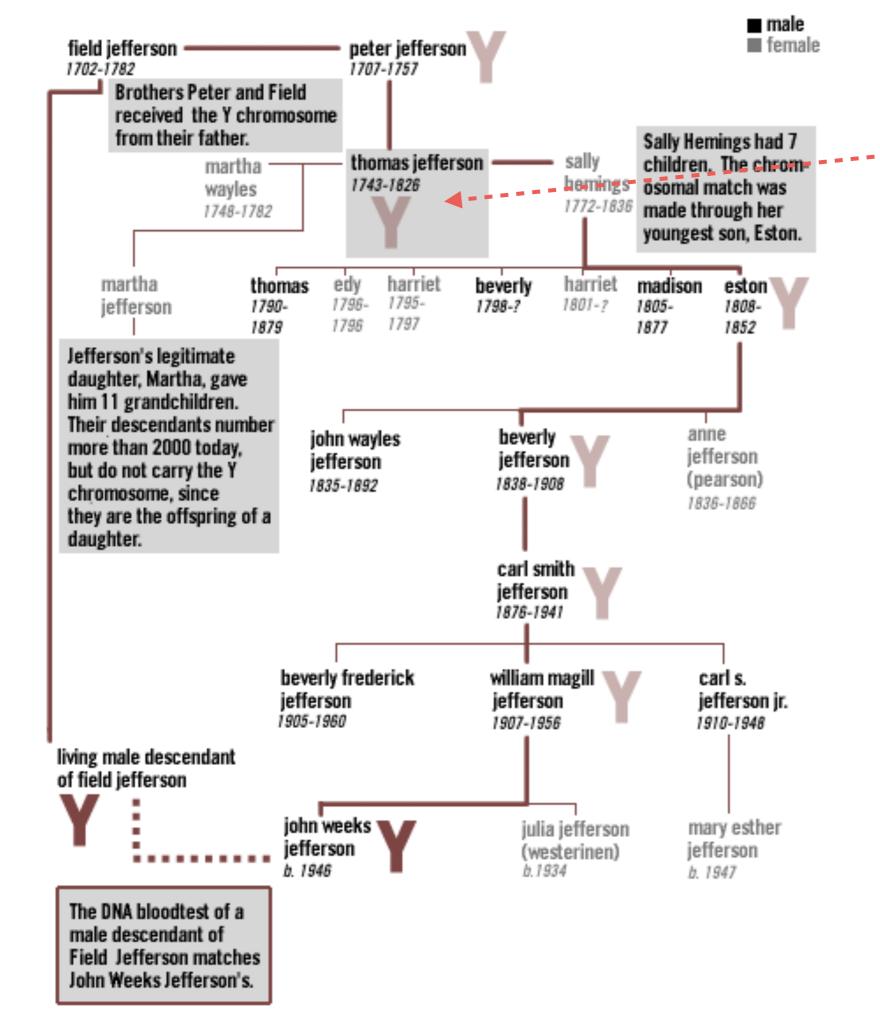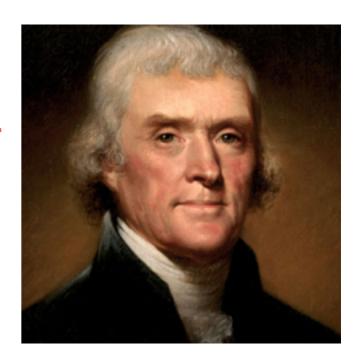# Forensic DNA analysis of historical samples

Two case studies

Thomas Jefferson and Sally Hemings

http://www.pbs.org/wgbh/pages/frontline/shows/jefferson/true/
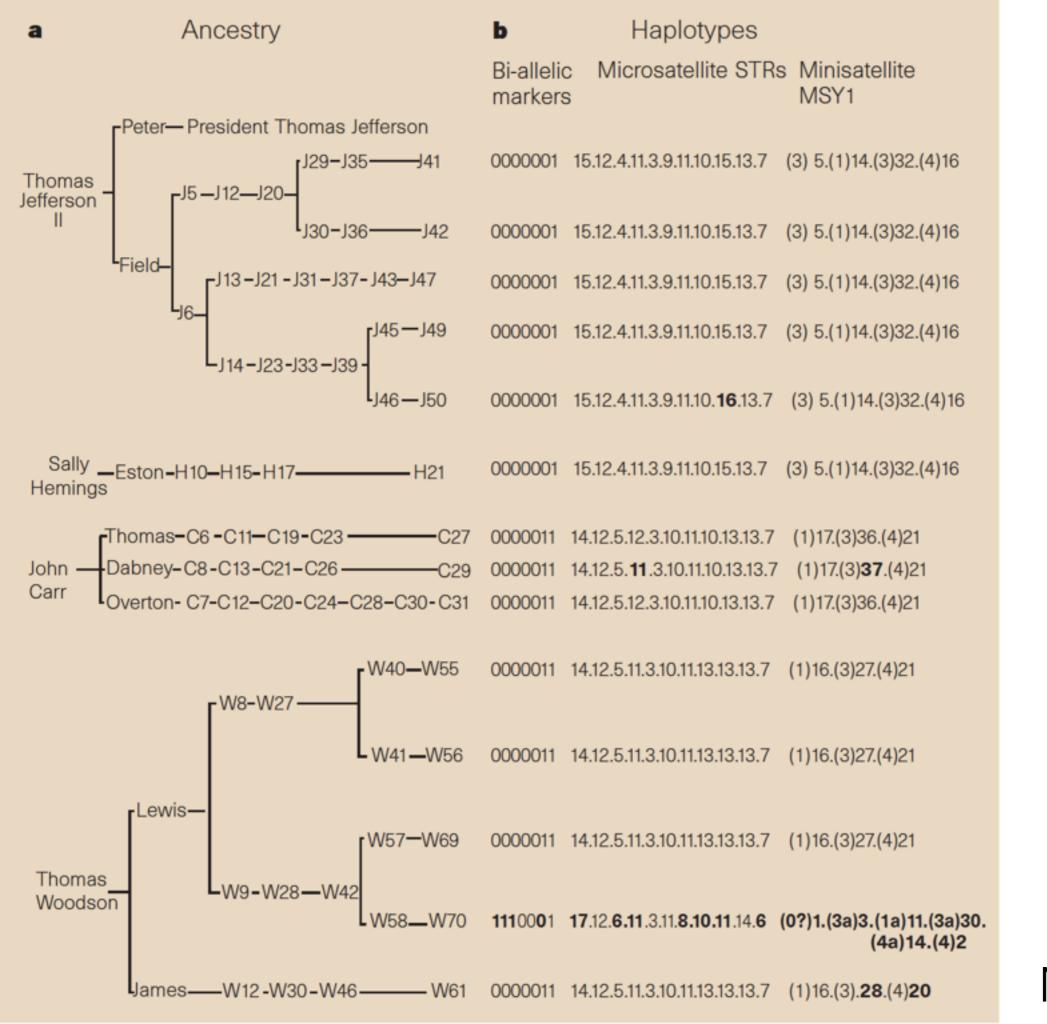
■ male
■ female

field jefferson
1702-1782

peter jefferson
1707-1757

Brothers Peter and Field received the Y chromosome from their father.

martha wayles
1748-1782

thomas jefferson
1743-1826

sally hemings
1772-1836

Sally Hemings had 7 children. The chromosomal match was made through her youngest son, Eston.

martha jefferson

thomas
1790-1879

edy
1796-1796

harriet
1795-1797

beverly
1798-?

harriet
1801-?

madison
1805-1877

eston
1808-1852

Jefferson's legitimate daughter, Martha, gave him 11 grandchildren. Their descendants number more than 2000 today, but do not carry the Y chromosome, since they are the offspring of a daughter.

john wayles jefferson
1835-1892

beverly jefferson
1838-1908

anne jefferson (pearson)
1836-1866

carl smith jefferson
1876-1941

beverly frederick jefferson
1905-1960

william magill jefferson
1907-1956

carl s. jefferson jr.
1910-1948

living male descendant of field jefferson

john weeks jefferson
b. 1946

julia jefferson (westerinen)
b.1934

mary esther jefferson
b. 1947

The DNA bloodtest of a male descendant of Field Jefferson matches John Weeks Jefferson's.

a Ancestry   **b** Haplotypes

Bi-allelic markers	Microsatellite STRs	Minisatellite MSY1
0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
0000001	15.12.4.11.3.9.11.10.**16**.13.7	(3) 5.(1)14.(3)32.(4)16
0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
0000011	14.12.5.12.3.10.11.10.13.13.7	(1)17.(3)36.(4)21
0000011	14.12.5.**11**.3.10.11.10.13.13.7	(1)17.(3)**37**.(4)21
0000011	14.12.5.12.3.10.11.10.13.13.7	(1)17.(3)36.(4)21
0000011	14.12.5.11.3.10.11.13.13.13.7	(1)16.(3)27.(4)21
0000011	14.12.5.11.3.10.11.13.13.13.7	(1)16.(3)27.(4)21
0000011	14.12.5.11.3.10.11.13.13.13.7	(1)16.(3)27.(4)21
1110001	**17**.12.**6.11**.3.11.**8.10.11**.14.**6**	(0?)1.(3a)3.(1a)11.(3a)30.(4a)14.(4)2
0000011	14.12.5.11.3.10.11.13.13.13.7	(1)16.(3).**28**.(4)**20**

Foster et al
Nature, 1998

a Ancestry   **b** Haplotypes

	Bi-allelic markers	Microsatellite STRs	Minisatellite MSY1
J41	0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
J42	0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
J47	0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
J49	0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
J50	0000001	15.12.4.11.3.9.11.10.**16**.13.7	(3) 5.(1)14.(3)32.(4)16
H21	0000001	15.12.4.11.3.9.11.10.15.13.7	(3) 5.(1)14.(3)32.(4)16
C27	0000011	14.12.5.12.3.10.11.10.13.13.7	(1)17.(3)36.(4)21
C29	0000011	14.12.5.**11**.3.10.11.10.13.13.7	(1)17.(3)**37**.(4)21
C31	0000011	14.12.5.12.3.10.11.10.13.13.7	(1)17.(3)36.(4)21
W55	0000011	14.12.5.11.3.10.11.13.13.13.7	(1)16.(3)27.(4)21
W56	0000011	14.12.5.11.3.10.11.13.13.13.7	(1)16.(3)27.(4)21
W69	0000011	14.12.5.11.3.10.11.13.13.13.7	(1)16.(3)27.(4)21
W70	**1110001**	**17**.12.**6**.**11**.3.11.**8**.**10**.**11**.14.**6**	**(0?)**1.(3a)3.(1a)11.(3a)30.(4a)14.(4)2
W61	0000011	14.12.5.11.3.10.11.13.13.13.7	(1)16.(3).**28**.(4)**20**

Foster et al
Nature, 1998

- *'TJF and most historians believe that, years after his wife's death, Thomas Jefferson was the father of the six children of Sally Hemings mentioned in Jefferson's records, including Beverly, Harriet, Madison, and Eston Hemings.'* — Thomas Jefferson Foundation (TJF), https://www.monticello.org/site/plantation-and-slavery/thomas-jefferson-and-sally-hemings-brief-account

- But the Thomas Jefferson Heritage Society disputes this conclusion

- http://uvamagazine.org/articles/anatomy_of_a_mystery/

The Romanovs

Published study in 1994 showed remains belonged to Romanov family

Table 1 STR genotypes[a] for the nine skeletons

Skeleton	HUMVWA/31	HUMTH01	HUMF13A1	HUMFES/FPS	HUMACTBP2
1	14,20	9,10	6,16	10,11	ND
2	17,17	6,10	5,7	10,11	11,30
3	15,16	8,10	5,7	12,13	11,32
4	15,16	7,10	7,7	12,12	11,32
5	15,16	7,8	5,7	12,13	11,36
6	15,16	8,10	3,7	12,13	32,36
7	15,16	8,8	3,5	12,13	32,36
8	15,17	6,9	5,7	8,10	ND
9	16,17	6,6	6,7	11,12	ND

Gill et al, Nature Genetics 1994
http://www.nature.com/ng/journal/v6/n2/abs/ng0294-130.html

Table 1 STR genotypes[a] for the nine skeletons

Skeleton	HUMVWA/31	HUMTH01	HUMF13A1	HUMFES/FPS	HUMACTBP2
1 (servant)	14,20	9,10	6,16	10,11	ND
2 (doctor)	17,17	6,10	5,7	10,11	11,30
3 (child)	15,16	8,10	5,7	12,13	11,32
4 (Tsar)	15,16	7,10	7,7	12,12	11,32
5 (child)	15,16	7,8	5,7	12,13	11,36
6 (child)	15,16	8,10	3,7	12,13	32,36
7 (Tsarina)	15,16	8,8	3,5	12,13	32,36
8 (servant)	15,17	6,9	5,7	8,10	ND
9 (servant)	16,17	6,6	6,7	11,12	ND

Gill et al, Nature Genetics 1994
http://www.nature.com/ng/journal/v6/n2/abs/ng0294-130.html

Table 2 Summary of mtDNA differences compared to the Anderson[21] reference sequence

Origin of sample	DNA source	Length sequenced (bp)	Positions within hypervariable regions (HVR) of mitochondrial DNA																			
			HVR 1												HVR 2						
			16111	16126	16169	16261	16264	16278	16293	16294	16296	16304	16311	16357	73	146	195	263	309.1	309.2	315.1
			C	T	C	C	C	C	A	C	C	T	T	T	A	T	T	A	*	*	*
	Femur skeleton 1	760	.	.	.	.	.	.	.	.	.	C	.	.	–	.	.	G	C	C	C
	Femur skeleton 8	742	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	C
	Femur skeleton 9	650	.	.	.	T	.	T	G	.	.	.	C	.	–	.	C	G	.	.	C
	Femur skeleton 2	736	.	.	.	.	T	.	.	.	.	.	.	.	.	C	.	G	C	C	C
	Femur skeleton 3	755	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
	Femur skeleton 5	634	T	.	.	.	.	.	.	.	.	.	.	C	.	.	–	G	.	.	C
	Femur skeleton 6	760	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
	Femur skeleton 7	744	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
	Blood sample	760	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
	Femur skeleton 4	782	.	C	Y	.	.	.	.	T	T	.	.	.	G	.	.	G	.	.	C
	Blood sample	781	.	C	T	.	.	.	.	T	T	.	.	.	G	.	.	G	.	.	C
	Blood sample	782	.	C	T	.	.	.	.	T	T	.	.	.	G	.	.	G	.	.	C

., Sequence unchanged from reference sequence; –, No nucleotide assignment; *, Nucleotide absent from reference sequence; Y, C/T heteroplasmy.

Gill et al, Nature Genetics 1994
http://www.nature.com/ng/journal/v6/n2/abs/ng0294-130.html

Table 2 Summary of mtDNA differences compared to the Anderson[21] reference sequence

Origin of sample	DNA source	Length sequenced (bp)	16111	16126	16169	16261	16264	16278	16293	16294	16296	16304	16311	16357	73	146	195	263	309.1	309.2	315.1
			C	T	C	C	C	C	A	C	C	T	T	T	A	T	T	A	*	*	*
Servant 1 (?)	Femur skeleton 1	760	.	.	.	.	?	.	.	.	.	C	.	.	−	.	.	G	C	C	C
Servant 2 (?)	Femur skeleton 8	742	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	C
Servant 3 (?)	Femur skeleton 9	650	.	.	.	.	T	.	T	G	.	.	.	C	−	.	C	G	.	.	C
Royal Physician Dr Botkin (?)	Femur skeleton 2	736	.	.	.	.	T	.	.	.	.	.	.	.	.	C	.	G	C	C	C
Daughter 1 of Tsar/Tsarina (?)	Femur skeleton 3	755	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
Daughter 2 of Tsar/Tsarina (?)	Femur skeleton 5	634	T	.	.	.	.	.	.	.	.	.	.	C	.	.	−	G	.	.	C
Daughter 3 of Tsar/Tsarina (?)	Femur skeleton 6	760	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
Tsarina Alexandra (?)	Femur skeleton 7	744	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
Duke of Edinburgh (Grand nephew of Tsarina)	Blood sample	760	T	.	.	.	.	.	.	.	.	.	.	C	.	.	.	G	.	.	C
Tsar Nicholas II (?)	Femur skeleton 4	782	.	C	Y	.	.	.	.	T	T	.	.	.	G	.	.	G	.	.	C
Gt. Gt. grandson of Louise of Hesse-Cassel	Blood sample	781	.	C	T	.	.	.	.	T	T	.	.	.	G	.	.	G	.	.	C
Gt. Gt. Gt. granddaughter of Louise of Hesse-Cassel	Blood sample	782	.	C	T	.	.	.	.	T	T	.	.	.	G	.	.	G	.	.	C

., Sequence unchanged from reference sequence; −, No nucleotide assignment; *, Nucleotide absent from reference sequence; Y, C/T heteroplasmy.

Gill et al, Nature Genetics 1994
http://www.nature.com/ng/journal/v6/n2/abs/ng0294-130.html

What about the other two children?

	VWA	TH01	F13A1	FES/FPS	ACTBP2	AMELOGENIN
Tsar (Skeleton 4)[a]	15,16	7,9,3[b]	7,7	12,12	11,32	X,Y
Tsarina (Skeleton 7)[a]	15,16	8,8	3.2[c],5	12,13	32,36	X,X
Anna Anderson (intestine sample)	14,16	7,9.3	3.2,7	11,12	15,18	X,X

Table 2 Mitochondrial DNA sequences

Origin of sample	DNA source	Length sequenced (bp)	Position within the non-coding region (ref. 11)					
			16111	16126	16266	16294	16304	16357
Reference sequence	–	–	C	T	C	C	T	T
Duke of Edinburgh (Great nephew of Tsarina)	Blood sample	403	T	.	.	.	.	C
Anna Anderson	Intestine sample	403	.	C	T	T	C	.
Anna Anderson	Hair sample	344–362 (3 hairs)	.	C	T	T	C	.
C. Maucher (Great nephew of Schanzkowska	Blood sample	380	.	C	T	T	C	.

., Sequence unchanged from reference sequence.

Stoneking et al, Nature Genetics 1995

Remains of two other bodies discovered 2007

Table 3. Autosomal STR Genotypes for the Romanov Family.

Marker	Sample 4.3	Sample 7.4	Sample 3.46	Sample 5.21	Sample 6.14	Sample 147	Sample 146.1
	Tsar Nicholas II	Tsarina Alexandra	Olga	Tatiana	Maria or Anastasia	Anastasia or Maria	Alexei
Amelog	X, Y	X, X	X, X	X, X	X, X	X, X	X, Y
D3S1358	14, 17	16, 18	17, 18	17, 18	16, 17	17, 18	14, 18
TH01	7, 9.3	8, 8	8, 9.3	7, 8	8, 9.3	7, 8	8, 9.3
D21S11	32.2, 33.2	30, 32.2	30, 33.2	32.2, 33.2	30, 33.2	30, 33.2	32.2, 33.2
D18S51	12, 17	12, 13	12, 12	12, 12	13, 17	12, 17	12, 17
D5S818	12, 12	12, 12	12, 12	12, 12	12, 12	12, 12	12, 12
D13S317	11, 12	11, 11	11, 11	11, 11	11, 11	11, 11	11, 12
D7S820	12, 12	10, 12	12, 12	10, 12	12, 12	10, 12	12, 12
D16S539	11, 14	9, 11	11, 11	11, 11	11, 14	9, 11	11, 14
CSF1PO	10, 12	11, 12	11, 12	11, 12	10, 11	10, 12	10, 12
D2S1338	17, 25	19, 23	17, 19	23, 25	17, 19	17, 23	23, 25
vWA	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16	15, 16
D8S1179	13, 15	16, 16	13, 16	15, 16	13, 16	15, 16	15, 16
TPOX	8, 8	8, 8	8, 8	8, 8	8, 8	8, 8	8, 8
FGA	20, 22	20, 20	20, 22	20, 20	20, 22	20, 22	20, 22
D19S433	13, 13.2	13, 16.2	13.2, 16.2	13.2, 16.2	13, 16.2	13, 13	13, 13.2

Topics	Data type	Information wanted	Main tool
HIV	nucleotide sequences (DNA)	evolutionary relationship	distances, visualizations
Personal Genomics	single nucleotide polymorhisms (SNPs)	trait association	hypothesis tests
Forensics	short tandem repeats (STRs)	identity	probability

Take-away messages

1. getting data ready for analysis can be time-consuming for non-routine data types

2. **making (reasonable) assumptions is difficult**

3. programming automates the calculations, making this step easy

4. **interpreting results properly is difficult**