Complete Genealogies

Reagan W. Moore

rwm@coregen.center

P. O. Box 3026

Chapel Hill, NC 27515-3026

May 2, 2020

I. Introduction

Every genealogy has properties related to completeness, correctness, consistency, and closure. This paper examines metrics that can be used to evaluate each property, and illustrates the results through application to a Research Genealogy of 300,000 persons. These generic properties can then be used to evaluate the attributes of a complete genealogy, including identification of the persons with the highest connectivity and progenitor persons. These attributes are used in turn to define the concept of a Core Genealogy, the set of persons to which all members of a cultural group can link their ancestry. The viability of a Core Genealogy is demonstrated by testing whether significant groups of persons can be linked, such as the United States presidents and the historically notable persons listed in Palmer's "History of the Modern World". The result of these analyses is the creation of a Core Genealogy for persons of Western European descent, based on the common ancestors of the Royal Lines of Europe. The Core Genealogy then provides a genealogical history of the modern world, linking historically notable persons to a common genealogy.

The Research Genealogy that is used to evaluate these concepts was assembled by adding 50 names per night for 25 years. More than 1400 sources were accessed, including history books, compiled genealogies, census data, web sites, and family records. This generated 450,000 source references for the 300,000 persons. Once the size of the Research Genealogy reached 5,000 persons, a corresponding analysis program was developed iteratively over 15 years. Each time a new genealogical property was considered, the Research Genealogy was revised to provide the required information for the analysis. The revisions were driven by the desire to remove errors (improve correctness), ensure information uniformity for all persons (consistency), eliminate disconnected groups (closure), and decide when the Research Genealogy was complete (completeness). These properties can be evaluated for any genealogy using the free CoreGen3 program available from the Mac App Store (SM).

This paper describes the development basis for the metrics for each genealogy property. The methods used in the CoreGen3 program for implementing each metric are described. The results of applying the metric to the Research Genealogy are then provided. The hope is that similar analyses will be done by each web site that provides genealogy information, with the goal of eventually agreeing on a Core Genealogy for persons of Western European Descent.

II. Correctness

The trustworthiness of a genealogy is strongly tied to the trustworthiness of the sources. Unfortunately, virtually every source used in the construction of the Research Genealogy contained errors. These ranged from incorrect dates for events, to incorrect relationships, to incomplete information about place locations and occupations or titles. Initially two methods were used to identify poor information sources.

1. The internal consistency of the information in the source was tracked. Typically each source provided information about a person multiple times: as a parent, as a spouse, and as a child. If the information was inconsistent within the source, the source was considered less trustworthy and information from other sources was rated higher.
2. The consistency of information between sources was tracked. The goal was to find the same information in at least two unrelated sources, preferably from different cultural groups. This was not always possible. If the sources disagreed, the more detailed information was usually considered to be more trustworthy. In the Research Genealogy each person has between 1 and 41 sources, with an average of 1.5 sources per person.

A metric for correctness is needed that is independent of the source. A property that applies to every person is biological feasibility. For each event in a person's life, a range of biologically feasible ages can be assigned for men and women. Persons that have ages that fall outside these ranges are suspect and candidates for review. This requires that every person in the genealogy have an estimated birth year. The ages at marriage, birth of a child, and death can then be evaluated and compared to the biologically feasible age ranges.

In the CoreGen3 program, the "Error #" operation within the Root Operation window evaluates correctness of all dates.

II.1 Correctness Metric

The metric for correctness applied in the CoreGen3 program checks age ranges for every person. The age limits were chosen broadly to be consistent with known events. Persons that are not compliant are annotated with the first detected error condition. Summary statistics are generated separately for men and women for distributions by century of the age at marriage, age at birth of a child, and age at death. The error metrics are defined below:

Men:

- Check whether age at marriage is more than 55 years older than wife
- Check whether age at birth of child is between 11 and 80
- Check whether husband is not male

Women:

- Check whether age at marriage is more than 30 years older than husband
- Check whether age at birth of child is between 11 and 55
- Check whether wife is not female

Men and Women:

- Check whether age at death is > 120 years
- Check whether death year is before birth year
- Check whether person born after parent's death
- Check whether person married before birth
- Check whether person married after death
- Check whether spouse is indicated by marriage date but name is missing
- Check whether a child family is missing
- Check whether gender is not known
- Check whether birth year is missing

II.2 Research Genealogy Correctness

The construction of the Research Genealogy was strongly limited by the availability of information as lineages are traced into the past. Fortunately, the lineages that are most well known are the Royal Lines of Europe for the Kings and Queens of Spain, Great Britain, Belgium, Netherlands, Denmark, Sweden, and Norway. These lineages include not only historical facts, but also links to cultural tradition. The estimation of a birth year is essential for building coherent time lines between the cultural groups. It was possible to estimate biologically feasible dates for all cultural groups except for the Davidic King list. The traditional dates proposed by James Ussher, Archbishop of Armagh, were used instead. This means that the Research Genealogy contains by default, 87 known biological inconsistencies for persons that lived too long or had children at too old an age, or had too large an age difference between spouses. The Royal Lineages provided links back to the following cultural groups in Europe:

- Anglo Saxons - Kings of Essex, Sussex, Wessex, Mercia, Northumberland, and Kent
- Asturia - Imams of Seville
- Carolingians - French descendants of Charlemagne
- Davidic King list back to Adam and Eve
- Eastern Roman Emperors of Byzantium
- Egyptian Pharaohs
- English peerage to the 1800s
- Franks - descendants of Clovis I
- German peerage
- Irish Monarchs - High Kings of Northern Ui Neill, Southern Ui Neill, Leinster, Meath(Tara), Munster(Cashel), Ulster(Aileach), Breifni, Tyrone(Ulidia, Oriel), Dublin
- Latin Kings of Jerusalem during the Crusades
- Lombards - Italian and French lineages
- Macedonians - Cousins of Alexander the Great
- Merovingians - Frankish descendants of Merovech
- Ostrogoths - King Theodoric the Great
- Persians - Darius the Great
- Ptolemaic rulers of Egypt down to Cleopatra
- Romans - Emperors back to Julius Caesar
- Russians - Dukes of Kiev and the Tsars of Russia
- Scandinavians - Kings back to Odin
- Scots - Kings of Dalriada
- Spanish - Kings of Aragon, Castile, Leon, Navarre
- Vandals - Kings Geiseric, Hunneric, and Hilderic
- Visigoths - Kings Theodoric, Alaric, and Athanagild
- Welsh - Kings of Gwynedd, Deheubarth, and Powys

In Table 1., distributions by century of the ages at marriage, age at birth of a child, and age at death are provided for both men and women. While an estimated birth year was created for every person, not all persons had information about marriage dates or death dates. The average age at birth of a child for men is 32 years per generation and for women is 28 years per generation. Note this average is taken across all children births, not just the eldest. The average age at death has slowly increased from 50 to 60. Note that limited statistics are available for persons in the 21st century. In Table 1,

Gen - average age of parent at birth of all children within century

Age - average age at death for all persons born within century

Mar - average age of person at marriage for marriages in century

Table 1. Average Ages

Year	male-gen	female-gen	male-age	female-age	male-mar	female-mar
000	33	27	73	48	31	20
100	30	24	50	45	19	14
200	29	24	55	53	28	29
300	29	25	51	43	35	23
400	29	26	50	54	29	24
500	29	27	50	42	28	24
600	30	25	49	50	25	21
700	29	25	51	45	29	20
800	29	25	49	50	29	22
900	30	25	49	52	29	22
1000	31	26	50	50	29	22
1100	31	26	51	51	29	22
1200	31	26	49	49	28	21
1300	31	26	50	50	28	22
1400	31	27	53	52	29	23
1500	32	27	53	53	29	23
1600	33	28	55	54	29	23
1700	33	29	58	56	28	23
1800	34	28	60	62	30	24
1900	32	27	56	63	29	25
2000	35	33	0	0	0	0
Average	32	28	55	55	29	23

In Table 2, a distribution by 5-year age interval is done for the number of men and women for age at marriage, age at birth of a child, and age at death.

Table 2. Distribution of number of persons per 5-year age interval

Age	male-gen	female-gen	male-age	female-age	male-mar	female-mar
0	0	0	3386	2434	14	21
5	0	0	678	427	54	113
10	108	368	563	359	479	1577
15	4138	12152	999	641	2897	10570
20	21799	47183	1937	1416	11859	15411
25	54167	53514	2699	2232	10959	7485
30	50815	38053	3673	2797	6347	3365
35	34109	22760	4453	3402	3556	1594
40	21701	8174	5327	3634	2225	751
45	11764	1680	6255	3576	1423	475
50	5614	292	6912	3779	905	274
55	2412	25	7613	4194	541	164
60	955	0	8112	4605	332	95
65	304	0	7850	4745	201	40
70	74	0	7567	4915	111	22
75	6	0	6531	4559	41	5
80	2	0	4922	3867	12	1
85	0	0	3090	2596	6	0
90	0	0	1417	1380	1	0
95	0	0	495	612	0	0
100	0	0	88	105	0	0
105	0	0	4	11	0	0
110	0	0	3	1	0	0
115	0	0	0	0	0	0
Total	207968	184201	84605	56289	41963	41963

Note that marriages may occur before the age of 11 in the form of marriage contracts in the Middle Ages. The Research Genealogy does contain 95 men and 117 women who lived past age 100.

Given that both historical facts and cultural tradition were integrated into the Research Genealogy, it was important to identify whether a source was more focused on cultural tradition than historical fact. To do this, a distribution by century of the number of persons mentioned by each source was constructed. Table 3 lists the distribution for the top 10 sources with the largest number of referenced persons. In the table,

Cent is the start of the century, followed by source numbers in the first row

Ave. is the birth year averaged across all referenced persons

Tot. is the total number of persons from the source summed across all referenced persons

Total is the total number of persons in the Research Genealogy born in the century

Information about persons born before 0 AD is included in the totals for the first century.

Table 3. Distribution of persons per century for the top 10 sources

Cent	711	589	260	224	258	216	1146	150	247	240	Total
0	128	0	0	0	1	0	0	0	0	0	1994
100	45	0	0	0	0	0	0	0	0	0	252
200	49	0	0	0	0	0	0	0	0	0	296
300	56	0	0	0	0	0	0	0	0	0	398
400	54	0	0	0	0	2	0	23	0	0	682
500	63	0	0	0	0	35	0	94	0	0	724
600	58	0	0	0	0	69	0	69	0	0	729
700	61	10	0	0	5	124	0	85	0	0	868
800	67	9	0	4	8	526	7	198	0	0	1697
900	148	0	0	41	34	836	15	312	0	0	2853
1000	619	0	23	254	100	1402	58	448	4	9	5415
1100	1413	0	398	782	256	1768	115	524	18	20	9611
1200	2251	6	1359	1998	494	1827	326	746	43	36	14503
1300	2887	270	3022	1791	978	1803	440	918	83	53	17594
1400	4256	1518	5496	1435	1976	1071	543	836	159	66	24716
1500	5784	4320	5138	1909	3162	425	1170	1361	1245	433	39034
1600	7838	8234	2437	1742	3139	350	3396	1803	4580	2749	62164
1700	7592	12615	54	1741	1544	458	2842	1169	2712	3134	61303
1800	5804	3350	0	2246	856	669	1801	1027	672	1795	37520
1900	1060	0	0	768	265	265	603	626	117	288	19887
2000	0	0	0	0	0	0	0	0	0	0	53
Ave.	1576	1684	1465	1545	1568	1286	1668	1497	1674	1725	1580
Tot.	40233	30332	17927	14711	12818	11630	11316	10239	9633	8583	302293

The source ID numbers in the top row of the table refer to the following sources:

- 711: Web Site, Geni, Geni, http://www.geni.com/
- 589: Web Site, Harrison, Bruce, The Family Forest Descendants of King Edward III of England and Queen Philippa of Hainault, http://www.familyforest.com
- 260: Book, Richardson, Douglas, Plantagenet Ancestry, Everingham, Kimball G., Genealogical Publishing Company

- 224: Book, 3.1 Isenburg, Wilhelm, Europaische Stammtafeln, Verlag von J. A. Stargardt, 1956, Vol 3.1, CS 616 I7 1956 v. 3
- 258: Book, Roberts, Gary Boyd, The Royal Descents of 600 Immigrants to the American Colonies or the United States, Genealogical Publishing Company, Incorporated
- 216: Book, 2 Schwennicke, Detlev, Europaische Stammtafeln, Stammtafeln zur Geschichte der Europaischen Staaten, Verlag von J. A. Stargardt, Vol 2
- 1146: Web Site, Dowling, Tim, Tim Dowlings Family Tree, http://gw.geneanet.org/tdowling
- 150: Book, 1.1, Schwennicke, Detlev, Europaische Stammtafeln, Neue Folge, Vittorio Klostermann, Frankfurt am Main, 1998, Vol I.1
- 247: Web Site, Lawson, Stephen M., Some Notable Cousins, http://kinnexions.com/kinnexions/cousins.htm
- 240: Web Site, Reitwiesner, William Addams, Ancestry of George W. Bush, http://www.wargs.com/political/bush.html

Sources with information about persons born before 1000 tend to include information from cultural tradition. The major sources for the Research Genealogy for information about Royal Lines were the 9 volumes of Collins' "Peerage of England" and the 24 volumes of Detlev Schwennicke's "Europaische Stammtafeln".

III. Consistency

Checking consistency requires the ability to sort information for each property to verify whether attributes are missing. The sorted list is then either searched or scrolled to check on individual values. This required the parsing of the semi-structured information available in a Gedcom file and the loading of the information into a graph database. A hierarchical, indexed, binary sort routine was developed to support sorting by strings, sub-strings, integers, and floating point numbers. In a hierarchical sort, if the values are the same in the first attribute, a check is made on the values of the second attribute, and so on. This makes it possible to examine consistency by sorting on last name or first name or person ID or any other attribute. A sort on person ID identifies holes in the genealogy where person IDs are missing.

The minimal information that is needed for each person to analyze properties of complete genealogies includes the following:

- Last name (maiden name for women)
- First name
- Person ID
- Relationship to another person
  - Parent Family ID (if the parents are known)
  - Spouse Family ID (if the spouse is known)
  - Child ID (if a child is known)
- Birth year
- Source ID for the information
- Country name included in each place location

When available, additional information was extracted from the Gedcom file, including prefixes, suffixes, titles, aliases, birth location, baptism location, marriage location and date, death location and date, and burial location. The information was loaded into a graph database to facilitate searching.

III.1 Research Genealogy Consistency

For the Research Genealogy, the desired information was not always available from each source. For instance, the last name may reflect the title or residence of the person. If the last name was missing for a women, the last name of the husband was entered in parentheses. If the first name was missing, a relationship designation was used, entered in parentheses, such as (daughter), (husband), (son). A birth year was estimated for every person based on the available information known about the families. A country name was entered for every place name.

Every person has source information. However for persons born after 1940, a privacy filter is applied that restricts the printed information to first name initials and birth year. No source information is listed for these persons. An exception is public information retrieved from Wikipedia and Wikitree about notable persons. All information known about notable persons is listed.

IV. Closure

The critical information in a genealogy is the relationships between persons. A genealogy has closure if it is possible to find a connection between any two persons in the genealogy. This means that there are no islands, with a group of persons isolated from the rest of the genealogy. To test for closure, the genealogy is partitioned into extended families. This process uses the following steps:

- 1- Set the partition number to 0
- 2- Select a root person
    - Find all of the relatives of the root person
    - Mark each person with the partition number, and save their relationship to the root
    - Make a list of all of the unrelated spouses
- 3- Increment the partition number
- 4- Loop over the list of unrelated spouses
    - For each unrelated spouse, find all of the relatives

- Mark each person with the partition number, and save their relationship to the unrelated spouse

- 5- Make a new list of unrelated spouses and iterate from step 3.

If the genealogy has closure, every person will have a partition number and a relationship to the root of their partition.

A connection between any two persons that traverses the minimum number of extended families can then be constructed by selecting one of the persons as root, partitioning the genealogy, and then tracing a lineage from the second person to the root of their partition, picking an unrelated spouse with the smallest partition number, and iterating. This will converge back to the person selected as root.

IV.1 Research Genealogy Extended Families

A partition of the 300,000 persons in the Research genealogy depends upon the selection of the person chosen as root. In table 4, an analysis of the extended families (partitions) in the Research Genealogy starting from Prince Georg Mountbatten-Windsor of Cambridge is shown. Degrees of separation are counted as the number of families that must be traversed to connect to the root partition. Each degree of separation can be thought of as another ring of families around the original root partition. In the table,

Partition-range lists the partition numbers in the ring of families
#Persons is the total number of persons in the ring of families
%Total is the cumulative percentage of persons across all inner rings

Table 4. Number of partitions in each ring of families

Degree	Partition-range	#Persons	%Total
0	1 - 1	163443	54.07
1	2 - 51294	112246	91.20
2	51295 - 62738	19536	97.66
3	62739 - 64043	3551	98.84
4	64044 - 64672	2105	99.53
5	64673 - 64986	915	99.84
6	64987 - 65132	280	99.93
7	65133 - 65170	99	99.96
8	65171 - 65191	65	99.98
9	65192 - 65205	32	99.99
10	65206 - 65214	18	100.00
11	65215 - 65217	3	100.00

There are 65,217 partitions, with one partition containing 163,443 persons. There are 49,330 partitions with a single person (no known relatives). The remaining partitions have between 2 and 558 persons.

IV.2 DNA distribution as a closure metric

An alternate way to define closure is to check whether an ancestral fractional distribution by birth country matches a biological DNA analysis. If a similar percentage of ancestral fraction distribution is assigned to each birth country, the genealogy adequately tracks the biological ancestry of the person. In the CoreGen3 program, clicking the AncDNA button in the Root Operations window automatically performs the following steps. Note that the steps can also be executed individually by clicking the appropriate buttons in the Root Operations window.

The steps involved in the analysis are:

- Construct an index for the birth countries (Country button in the Root Operations window).
- Identify the ancestors of the Root Person (Ancestors button in the Root Operations window).
- Calculate the ancestral fractional distribution by summing the distribution fraction for each child that is an ancestor of the root person. Each parent gets half of the sum.
- Identify the Treetops, the ancestors that have either 0 or 1 parent (TreeTops button in the Root Operations Window).
- For each Treetop, assign the appropriate fraction of the ancestral fractional distribution to the birth country (all if no parents, half if one parent). If the birth location is missing, the ancestral fraction is assigned to _unknown.

An example of the Ancestral Fractional Distribution for Prince George of Cambridge is given in Table 5, which identifies birth countries for 99.45% of his Ancestry.

Table 5. Distribution of Ancestral Distribution Fraction by birth country

Note that the birth country of the Treetops is used.
Number is the number of ancestors found in the country

Percent	Number	Country
64.9271%	4062	England
8.7155%	1166	Germany
8.2149%	649	Scotland
3.6541%	1308	France
3.2559%	251	Ireland
2.1132%	15	United States
1.1645%	61	Poland
1.0685%	254	Italy
0.6172%	90	Russia
0.5715%	114	Netherlands
99.4507%	9258	Total

The analysis is limited by the availability of information about ancestors. Thus 2.1% of the ancestral distribution fraction of Prince George is attributed to the United States, because the ancestry of those 15 persons could not be traced back to Europe.

V. Completeness

Popular completeness metrics include finding a royal descent, or a descent from Charlemagne. Each metric can be generalized. Instead of searching for a royal descent, a connection can be sought to the core genealogy for your cultural group. Once you connect, you will then be related to every other member of your cultural group who has linked their ancestry. In addition to a descent from Charlemagne, descents from the Progenitors of your cultural group can be identified. You will also have descents from hundreds of Kings.

A viable core genealogy needs the following properties:

- It must be highly interconnected. Connecting to one member of the core genealogy should enable relationships to all other members.
- It should extend from the present time back to persons defined in cultural traditions.
- It should be sufficiently large that connections can be made based on historical fact.
- It should provide links to as many cultural groups as possible.

A core genealogy based on the common ancestors of the Royal Lines of Europe satisfies all of these properties. The Royal Lines of Spain, Great Britain, Belgium, Netherlands, Denmark, Sweden and Norway are highly intermarried. Any descendant of one of these lineages can be chosen as the Root person for the core genealogy. The ancestors of the Root person then comprise a Core Genealogy for persons of Western European Descent. The core genealogy extends to the present time, and links to all of the cultural groups listed in section II.2.

The Core Genealogy analyzed in this paper is based on the ancestors of Prince George of Cambridge, the great-grandson of Queen Elizabeth of Great Britain. The Research Genealogy contains 36,000 ancestors for Prince George. In the CoreGen3 program, the high degree of intermarriage between the lines can be found by selecting Royal Lines and clicking on the button "Common Ancestors" in the Group operations list. Table 6 defines the relationships between the current Kings and Queens, using the following headers:

Rel is the relationship to the first person in the list
#Anc is the number of ancestors of the person
#CoreAnc is the number of core ancestors of the person

Table 6. Common Ancestors of the Royal Lines of Europe

	Person	Rel	#Anc	#CoreAnc
1	Philippe Leopold Louis Marie of Belgium	source	25651	22436
2	Margrethe II of Denmark	2C 1R	19788	19377
3	Elizabeth II Alexandra Mary Windsor	3C 1R	28091	28091
4	Willem Alexander of the Netherlands	4C 1R*	19746	19032
5	Harald V of Norway	1C 1R	19403	19268
6	Felipe VI of Spain	4C	22165	19669
7	Carl XVI Gustav of Sweden	3C	20905	19326

The total number of common ancestors is 18,805. Linking to the Core Genealogy provides linages to hundreds of kings. In the CoreGen3 program, clicking on the button Nobles in the Root Operations list generates summary statistics for the number of nobles present in the ancestors of Prince George (Root person), in the Core Genealogy (again Prince George), the descendants of Charlemagne, and in the Common Ancestors of the Royal Lines. In table 7,

#-Persons is the number of persons in the genealogy with each title.
Ancestors is the number of ancestors of the root person with each title.
Descendants is the number of descendants of the ancestral person with each title.
Coreanc is the number of core genealogy members with each title.
Comgroup is the number of common ancestors with each title

Table 7, Number of Nobles in Research Genealogy

Type	Title	#-Persons	Ancestors	Descendants	Coreanc	Comgroup
24	EMPEROR	230	96	94	96	96
23	MONARCH	178	100	0	100	89
22	PHARAOH	50	34	0	34	34
21	U.S. PRESIDENT	44	0	44	0	0
20	TSAR	62	11	26	11	5
19	SAINT	88	33	8	33	33
18	POPE	79	1	31	1	1
17	KING	2282	972	549	972	908
16	QUEEN	185	105	127	105	95
15	PRINCE	2744	286	1912	286	247
14	PRINCESS	1744	249	1502	249	198
13	DUKE	2712	657	2133	657	519
12	DUCHESS	684	166	621	166	109
11	KHALIF	70	0	1	0	0
10	EARL	2086	618	1756	618	220
9	COUNT	7900	2633	4287	2633	2467
8	COUNTESS	3389	843	2097	843	713
7	MARQUESS	850	85	476	85	53
5	MARQUESSA	136	38	94	38	28
4	BARON	3637	432	1436	432	152
3	BARONESS	782	70	206	70	25
2	LORD	4331	1677	2728	1677	578
1	KNIGHT	4318	1476	1754	1476	178

Note that while the Research Genealogy contains more than 2200 kings, more than 900 of the kings are present in the Core Genealogy (ancestors of Prince George). The Core Genealogy does contain Charlemagne, the ancestor of 549 Kings. Connections to the Core Genealogy automatically provide a descent from Charlemagne. The number of ascents can be found by clicking the "Ascents" button in the Root operations window when Prince George of Cambridge is selected as the Root person. In the Research Genealogy, Prince George has 1.6 billion ascents to Charlemagne. Each ascent takes a unique path through the genealogy.

A descendant fractional distribution from Charlemagne can be found by clicking on the "DescDNA" button in the Root Operations window when Charlemagne is selected as the Ancestral Person. Table 8 lists the person in each generation who has the largest number of descents. The high degree of connectivity is evident from the fact that a person born in the 1900's has a descendant fractional distribution from Charlemagne of 0.001. In the table,

```
Gen is the generation number
minDNA is the minimum expected descendant fraction assuming only one descent
DNA is the distribution fraction from Charlemagne across all descents
Birth is the Birth year
```

Table 8. Descendant fractional distribution from Charlemagne

Gen	minDNA	DNA	Birth	Person
1	5.00e-01	5.00e-01	773	King Pippin III of Italy
2	2.50e-01	2.50e-01	797	King Bernard de Vermandois
3	1.25e-01	1.88e-01	842	King Berengar I of Friuli
4	6.25e-02	1.25e-01	882	Gisela of Friuli
5	3.12e-02	7.81e-02	900	Berengar II of Ivrea
6	1.56e-02	5.08e-02	960	Susanne Rosala of Ivrea
7	7.81e-03	3.32e-02	959	Gisela of Burgundy
8	3.91e-03	3.00e-02	1020	Henry II of Brabant
9	1.95e-03	2.40e-02	1032	Matilda of Flanders
10	9.77e-04	1.71e-02	1088	Baldwin III of Hainault
11	4.88e-04	1.35e-02	1093	Baldwin VII of Flanders
12	2.44e-04	9.73e-03	1170	Isabelle of Hainault
13	1.22e-04	8.74e-03	1187	King Louis VIII of France
14	6.10e-05	7.48e-03	1215	King Louis IX of France
15	3.05e-05	6.81e-03	1248	Blanche d`Artois
16	1.53e-05	5.43e-03	1275	Ferdinand II de la Cerda
17	7.63e-06	5.07e-03	1312	Joan II of France
18	3.81e-06	4.89e-03	1332	King Charles II d`Evreux
19	1.91e-06	3.36e-03	1433	Francois II de Bretagne

20	9.54e-07	3.17e-03	1432	Royan de Bretagne
21	4.77e-07	2.80e-03	1519	King Henry II of France
22	2.38e-07	2.31e-03	1562	Charles Emmanuel I of Savoy
23	1.19e-07	2.18e-03	1545	Don Carlos of Spain
24	5.96e-08	1.92e-03	1605	Philipp of Austria
25	2.98e-08	1.86e-03	1669	Maria Antoinette Theresia Josepha of Austria
26	1.49e-08	1.61e-03	1710	King Louis XV of France
27	7.45e-09	1.44e-03	1751	King Ferdinand IV Antonia Pasquale of Naples and Sicily
28	3.73e-09	1.41e-03	1798	Marie Caroline Ferdinanda of Bourbon-Two Sicilies
29	1.86e-09	1.38e-03	1839	Karl Salvator of Habsburg-Tuscan
30	9.31e-10	1.34e-03	1862	Pedro de Alcantara de Bourbon y Bourbon
31	4.66e-10	1.29e-03	1903	Helena of Habsburg-Lorraine
32	2.33e-10	1.08e-03	1938	Carlos of Bourbon-Sicily
33	1.16e-10	6.51e-04	1994	S. of Wurttemberg

Because of the multiple descents, the actual descendant fractional distribution can be more than a million times larger than the minimal expected descendant fractional distribution.

V.1 Connectivity

The connectivity within the Core Genealogy can be analyzed locally by calculating the own-cousin relationship for each member. If a person has two ascents to an ancestor, the own-cousin relationship can be calculated by tracing the ancestry up one ascent and then the cousin relationship by going down the second ascent. The cousin relationship is then equal to the number of generations minus 1. If you have two ascents to your 4th Great-Grandfather, you are your own 5th cousin. In the CoreGen3 program, the own-cousin relationships for the Core Genealogy can be calculated by clicking on the "Core Own Cousin" button in the Core Operations window.. Table 9 provides a distribution of the number of persons with a given own cousin relationship by century for the core genealogy members. In Table 9,

```
Year is the start of the century.  Century 0 includes persons born before 0 AD
#OwnCos is the number of core genealogy persons who have an
  Own Cousin relationship
Additional columns are the # of persons with the given relationship
       1c - First Cousin
       2c - Second Cousin
       3c - Third Cousin, etc.
```

Table 9. Distribution of own cousin relationship by century

Year	#Owncos	1C	2C	3C	4C	5C	6C	7C	8C	9C	10C
0	394	14	41	35	33	21	16	16	13	11	9
100	95	0	4	7	3	4	7	9	6	5	4
200	99	0	2	1	6	10	10	2	4	5	6
300	129	0	1	5	4	3	3	12	16	19	13
400	165	1	9	7	9	6	11	13	4	7	20
500	174	0	4	17	22	22	11	6	5	5	2
600	165	0	1	4	10	24	22	19	11	6	3
700	215	0	8	15	19	15	24	11	20	15	15
800	390	1	8	31	60	68	40	22	18	23	21
900	679	1	15	49	92	133	116	93	48	24	14
1000	1212	4	10	59	182	228	221	181	105	67	39
1100	2099	3	32	108	218	405	391	292	246	133	84
1200	3017	0	10	100	419	686	542	388	255	187	127
1300	3091	2	16	124	470	645	533	371	283	182	125
1400	2703	0	14	137	417	582	428	322	253	174	109
1500	1817	0	14	106	351	391	286	201	134	100	83
1600	880	0	36	126	187	154	92	81	61	53	29
1700	267	0	14	46	49	34	21	22	18	16	9
1800	70	0	6	11	8	5	5	6	6	5	4
1900	12	0	0	1	3	4	2	0	0	1	0
2000	1	0	0	0	0	0	1	0	0	0	0
Total	17674	26	245	989	2562	3440	2782	2067	1506	1038	716

The peak of the distribution is 5th cousin which is consistent across the centuries. The implication is that typically the members of the Core Genealogy have two ascents to their 4th Great-Grandfather. When you connect a lineage to the Core Genealogy, within 6 generations links to other members of the Core Genealogy will emerge.

A remaining question is how far back in time a lineage must be traced to connect to the Core Genealogy. The connection year is analyzed in the CoreGen3 Program by clicking on the button "Connection Year" in the Core Operations window. For persons born in the 1900s, on average lineages connect to the Core Genealogy in 1625. For persons born in the 1800s, the connection year is 1583.

V.2 Progenitors

A set of Progenitors for persons of Western European descent can be generated by analyzing the global connectivity of the members of the Core Genealogy. The global connectivity is defined as the logarithm to the base 2 of the number of ascents to an ancestor, divided by the number of generations from the Root person to the ancestor. This metric approaches the value of 1 for highly interconnected genealogies.

In the CoreGen3 program, the global connectivity is calculated when you click on the "Core Own Cousin" button of the Core Operations window. In table 10,

```
Metric   = log2(# number of ascents from core root person) /
           (# of generations from core root person)
#gen     = number of generations from root of core genealogy to person
#Ascents = number of ascents from root of core genealogy to person
Birth    = Birth year
```

Table 10. Persons with the largest global connectivity values

	Person	Birth	#gen	#Ascents	Metric
1	Charlemagne the Great	747	33	1.66e+09	0.9281
2	Louis I the Pious	778	32	8.28e+08	0.9258
3	Hugh II de Tours	765	32	7.72e+08	0.9226
4	Aba de Morvois	779	32	7.72e+08	0.9226
5	Lothaire I of Italy	795	31	3.97e+08	0.9215
6	Ermengarde of Alsace	800	31	3.97e+08	0.9215
7	Rotrou of Austrasia	740	33	1.37e+09	0.9196
8	Girard I of Paris	735	33	1.37e+09	0.9196
9	Leutwinus of Treves	660	35	4.08e+09	0.9122
10	(daughter) of Neustria	662	35	4.08e+09	0.9122
11	Ermengarde of Hesbaye	778	32	5.61e+08	0.9083
12	Charles Martel	688	35	3.67e+09	0.9078
13	Rotrude of Alemania	688	35	3.63e+09	0.9074
14	Hildegarde of Swabia	757	33	1.02e+09	0.9071
15	Conrad I of Swabia	925	28	4.36e+07	0.9063
16	Yaroslav I Vladimirovi of Kiev	978	26	1.16e+07	0.9025
17	Ingegerd Olafsdotter	1001	26	1.16e+07	0.9025
18	Pippin II Quent of Vermandoise	817	30	1.40e+08	0.9021
19	Rothaeide di Bobbio	812	30	1.40e+08	0.9021
20	Sophia of Hungary	1046	25	6.10e+06	0.9017

The emergence of Charlemagne as the person with the highest global connectivity justifies the popular objective of seeking a descent from Charlemagne. However, there are many other persons with nearly the same degree of connectivity. Note that the persons come from France, Germany, Italy, Russia, Sweden, and Hungary. These persons are candidates for a set of Progenitors for persons of Western European Descent.

An alternate way to identify progenitors is to list the Tree Tops of the lineage coalescence points for the own cousin relationship. A lineage coalescence point is the first ancestor to which a person has two ascents. A Tree Top is a person who has a descendant that has two ascents, but does not themself have two ascents to any ancestor. For the Core Genealogy based on the ancestors of Prince George, there are 36,374 ancestors. Of these ancestors, 18,323 persons have an own cousin relationship. The number of Tree Tops is 530 persons. This means that lineages can be found for 18,323 persons to just 530 Progenitors. Every person of Western European descent will have lineages to a majority of the Progenitors in this list.

VI. Genealogical History of the Modern World

The Research Genealogy provides a genealogical history of the modern world, linking notable historical persons into a coherent genealogy. The genealogical history provides a unified time line for historically important persons across cultures and back to cultural traditions. The effectiveness of a Core Genealogy as the unifying component of a genealogical history of the modern world can be analyzed by examining whether important groups of persons can be linked.

VI.1 United States Presidents

A good test case is to examine whether the Presidents of the United States are related and linked to the Core Genealogy. In the CoreGen3 program, select the group marked U.S. Presidents, then click on "Common Ancestors" in the Group Operations window. This finds the relationship of each President to George Washington. The U.S. Presidents have 5981 common ancestors in the Research Genealogy. In Table 11,

```
Rel is the relationship to the first person in the list
#Anc is the number of ancestors of the person
#CoreAnc is the number of core ancestors of the person
The core ancestry is the ancestors of George Alexander Mountbatten-Windsor
```

Table 11. Common Ancestors of U. S. Presidents

	Person	Rel	#Anc	#CoreAnc
1	George Washington	source	13729	13104
2	John Adams	8C 3R*	12667	12497
3	Thomas Jefferson	9C 1R*	9797	9637
4	James Madison	10C 2R*	12153	11841
5	James Monroe	11C 3R*	10198	10103
6	John Quincy Adams	8C 4R*	12940	12606
7	Andrew Jackson	H 7C 3R*	11735	11649
8	Martin van Buren	11C 2R*	9542	9440
9	William Henry Harrison	10C	10902	10683
10	John IV Tyler	10C 3R*	10477	10314
11	James Knox Polk	11C 3R*	10028	9909
12	Zachary Taylor	9C	12420	12158
13	Millard Fillmore	10C 4R*	9917	9521
14	Franklin Pierce	9C 2R*	11519	11141
15	James Buchanan	11C 1R*	8259	8185
16	Abraham Lincoln	8C 4R*	13125	12726
17	Andrew Johnson	10C 4R*	9175	9061
18	Ulysses Simpson Grant	6C 2R*	13936	13425
19	Rutherford Birchard Hayes	7C 3R*	10830	10163
20	James Abram Garfield	11C 3R*	9439	8873
21	Chester Alan Arthur	10C 3R*	11045	10785
22	Stephen Grover Cleveland	7C 5R*	11444	11023
23	Benjamin Harrison	10C 2R*	11388	10966
24	Stephen Grover Cleveland	7C 5R*	11444	11023

25	William McKinley	8C 4R*	13246	13155
26	Theodore Roosevelt	8C 6R*	13471	13112
27	William Howard Taft	6C 5R*	15358	14383
28	Thomas Woodrow Wilson	11C 3R*	8195	8160
29	Warren Gamaliel Harding	8C 5R*	11567	11262
30	John Calvin Coolidge	6C 6R*	15354	13864
31	Herbert Clark Hoover	9C 7R*	12340	11720
32	Franklin Delano Roosevelt	8C 4R*	17240	16108
33	Harry S. Truman	9C 6R*	10724	10472
34	Dwight David Eisenhower	12C 5R*	8123	7945
35	John Fitzgerald Kennedy	11C 7R*	9413	9345
36	Lyndon Baines Johnson	9C 7R*	11368	11167
37	Richard Milhous Nixon	8C 7R*	13478	12903
38	Gerald Rudolph Ford	8C 8R*	14144	12941
39	James Earl Carter	H 9C 6R*	11494	10969
40	Ronald Wilson Reagan	8C 8R*	14270	14196
41	George Herbert Walker Bush	7C 8R*	17978	15874
42	William Jefferson Clinton	11C 7R*	9201	9097
43	George Walker Bush	7C 9R*	19515	16326
44	Barack Hussein Obama	3C 8R*	17662	16429
45	Donald John Trump	11C 8R*	10139	10014

The U. S. Presidents range from 3rd cousin to 12th cousin of George Washington, They are all connected to the Core Genealogy. The five most recent common ancestors of the U. S. Presidents are:

Person	ID	Birth	Death
1 Blanche d`Artois	6954	1248	1302
2 Eleanor of Castile	4107	1240	1290
3 Edward I Plantagenet	4104	1239	1307
4 Beatrice of Savoy	11721	1213	1258
5 Isabel Le Bigod	7328	1212	

Here ID is the person ID used in the Research Genealogy.

VI.2 Notable Historical Persons

A similar analysis can be performed for the persons listed in "The History of the Modern World" by R. Palmer and J. Colton. The index to the book lists 719 notable persons that have influenced historical events. A total of 600 of the persons are linked to the Research Genealogy, with 40% of the notables having enough information to define cousin relationships.

An ongoing project is to link as many of the additional historically important persons as possible to the Research Genealogy,. At that point, determination of the genealogical relationship between the persons that influenced history will be possible.

VII Conclusion

The Research Genealogy is being continually updated. Current goals include linking more notable historical persons, linking noteworthy persons alive today, comparing with other published Royal Lineages, and finding additional metrics for evaluating properties of genealogies. A recent extension is a comparison of the ancestors of King Charles II listed in the Research Genealogy with the lineage published in "The Ancestry of Charles II, King of England, A Medieval Heritage", compiled by Charles M. Hansen and Neil D. Thompson. Since they used the Europaische Stammtafeln as their principal source, the lineages compared almost exactly. The Research Genealogy is published at https://gw.geneanet.org/rwmoore_w?lang=en. An Apple version of the CoreGen3 genealogy analysis workbench is freely available at the Mac App Store (SM). A Windows version of the workbench had been developed and will be freely published.