

Lena, Maksim, Karina, bedankt voor de ondersteuning

Ir. Erik Mols

December 2021, eerste editie

Index

Inleiding	7
Het Filin Genetic Genealogy Raamwerk	13
Knowledge	17
Biology (Biologie)	18
De cel, bouwsteen van het leven	19
De Cel	20
Celkern (Nucleus)	22
Chromosomen	23
Mitochondria	26
DNA	28
DNA kopieer proces	31
Introns en exons	36
Evolutie de basis van Genetische Genealogie	39
SNP	45
Indel	47
STR	48
GCN	50
Wetten en Regels	51
Software	53
Archieven	55
Digital Resources (digitale bronnen)	56
Bronnen voor Nederland en België:	57

Een waarschuwing voor onderzoekers	58
Historische Context	59
Ethica en genealogie	61
Niet gewenste resultaten	62
NPE	63
Extract and Load(Extraheren en Laden)	64
Mapping our DNA (DNA in kaart brengen)	65
Chip testing (Chip testen)	66
Whole Genome Sequencing (WGS)	70
Starting with the Extracting phase (Starten met extraheren)	74
De verschillende test bedrijven	75
Ancestry	76
23andMe	77
MyHeritage	78
LivingDNA	79
FTDNA	80
Andere Genetische Genealogie bedrijven	80
Research (Onderzoek)	82
Research Questions (Onderzoek Vragen)	83
Methodology (Methodologie)	85
ACT (handelen)	86
Results (Resultaten)	86

cM	87
Autosomal results (autosomale resultaten)	93
Haplogroup (haplogroep) en familienamen	101
Health results (Gezondheid resultaten)	112
aDNA of Ancient DNA (Antiek DNA)	113
Evaluate (Evalueren)	116
Data management	118
Filestructures (Bestandsstructuren)	118
Logs	120
Backup	121
Exchange (Uitwisselen)	122
Accounts	125
Clustering and tree building (Groeperen en stambomen maken)	127
Native tree (oorspronkelijke stamboom)	128
Bettinger cM tool	132
Draft tree (stamboom schets)	135
WATO	135
Separate Paternal Maternal (scheiden van ..)	140
Autoclustering	140
DNApainter	142
Starting to paint (Starten met verven)	145
Data verkrijgen van myHeritage	151

Data verkrijgen van Ancestry	152
Data verkrijgen van 23AndMe	152
Data verkrijgen van LivingDNA	154
Data verkrijgen van Gedmatch	154
Chromosome Browser	155
Leeds	159
Phasing, Lazarus and Evil Twin	168
Evaluate (Evalueren)	171
yDNA mtDNA results	172
Clustering Resultaten	178
Ethnicity Archeo Admixture	178
Traditional Genealogy (Traditionele Genealogie)	179
Answer Research Question (Beantwoorden van de onderzoeksvragen)	179
Social Results (Sociale Resultaten)	179
Evaluate Tree (Stamboom evalueren)	180
Traditional Genealogy (Klassieke Genealogie)	184
Voor 1545	187
Na 1545	190
Na 1796	190
Heraldiek	192
Kerk Archieven	196
Juridisch	197

Gemeente Administratie	198
Census	198
Folklore	199
Social Research	200
Social Media	203
Traditionele Media	203
Google Earth	205
Archeogenetics, Admixture en Etniciteit	206
Pool based algorithms	208
Archeogenetics	209
Mijn persoonlijk verhaal	216
Index	221

Inleiding

Voordat ik daadwerkelijk ga starten met het uitleggen van Genetica en Genetische Genealogie, wil ik eerst mijzelf introduceren. Mijn naam is Erik Mols. Ik ben geboren in Tilburg in het prachtige jaar 1970. Na de middelbare school heb ik biologie en bio-informatica gestudeerd. Ik heb een bachelor graad in Biologie van “De hogeschool katholieke leergangen” en een Mastergraad in bio-informatica van de “Wageningen Universiteit”. Momenteel ben ik actief als zelfstandig ondernemer, waarbij ik met name werk voor commerciële Hogescholen en Universiteiten. In het hoger onderwijs ben ik actief als onderwijsontwikkelaar, auteur, docent en examinator. De onderwerpen waarin ik lesgeef zijn voor een groot gedeelte Informatica gerelateerd, maar ik geef ook les in Data Science, Moleculaire Biologie en Wiskunde. Tijdens mijn studie en gedurende mijn carrière heb ik meegewerkt in diverse wetenschappelijke projecten, welke meestal niet gepubliceerd werden. In hbo-land is het gebruikelijk dat het

onderzoek op het tweede plan komt. Onderzoeksprojecten van Lectoraten worden zelden gepubliceerd, mede ook omdat ze vaak in opdracht van het bedrijfsleven worden uitgevoerd.

Ik ben getrouwd met Lena en samen hebben we twee kinderen, Maksim en Karina. Wij wonen in het prachtige dorp Biezenmortel, wat tot de gemeente Tilburg hoort.

Dit boek is opgezet als een werkboek. Het is gecentreerd rond het Genetische Genealogy Raamwerk, wat ik gecreëerd heb. Ik heb ervoor gekozen geen tweede editie uit te brengen van mijn boek: “An unbiased Genetic (DNA) Genealogy textbook”, maar een nieuw boek te schrijven, wat gefocust is rondom het raamwerk. In plaats van alleen maar informatie bevat het een methode om Genetische Genealogie Puzzels op te lossen. Naast deze methodologie bevat het boek een revisie van veel van mijn teksten uit “An unbiased Genetic (DNA) Genealogy Textbook”, maar het bevat ook veel totaal nieuwe teksten. Dit werkboek is niet alleen

geschreven als een introductie voor mensen die met genetische genealogie willen starten, maar het is zeker ook geschikt voor mensen die er al wat langer mee bezig zijn. Door de teksten te schrijven rondom het raamwerk is het een werkboek geworden. Ook probeer ik in het werkboek genetische en traditionele genealogie met elkaar te verbinden, omdat technieken uit beide werelden elkaar versterken. Ik vind het belangrijk om te vermelden dat ik niet verbonden ben aan een van de vele genetische genealogie bedrijven, derhalve is mijn boek onafhankelijk en grotendeels objectief. Grotendeels objectief, omdat ik natuurlijk wel mijn eigen voorkeuren heb.

Maar laat ik eerst eens beginnen met een goede academische discussie over het begrip Genetische Genealogie. Sommige lezers die mij al kennen, weten dat ik de term genetische genealogie verafschuw. Laat ik mij nader verklaren. Tijdens mijn studie is mij geleerd, dat Genetica een erg breed vakgebied is. Het bestaat uit allerlei omics: Genomics, Transcriptomics, Proteomics en Metabolomics.

Wanneer je kijkt naar het totale vakgebied, maken we binnen de Genetische Genealogie eigenlijk alleen gebruik van Genomics. Het is mij dan ook volstrekt onduidelijk waarom de term Genetische Genealogie ooit gekozen is. Genomische Genealogie zou een benaming zijn die de lading veel beter dekt. Of wat te denken van DNA Genealogie. Immers, behalve de data uit onze Genomen, worden er geen andere onderdelen van de Genetica gebruikt. Om de genetische genealogie begrippen en technieken uit te leggen, moet ik soms simplificeren, ik doe dit om dit boek voor iedereen toegankelijk te maken.

Voordat ik diep in het raamwerk en de biologie ga duiken wil ik eerst een aantal basisbegrippen verkennen, welke nodig zijn om het genealogische deel van de Genetische Genealogie te begrijpen. Het eerste begrip is maternal, wat afstamt van het Latijnse woord mater. Mater vertaald naar het Nederlands is moeder. Het equivalent van Mater is Pater, wat voor Vader staat. Het begrip paternal is van pater afgeleid. Nal, zouden we kunnen vertalen

als kant. Wat vervolgens betekent dat maternal staat voor moederskant en paternal voor vaderskant.

In Genealogie kennen we het begrip 'Most Recent Common Ancestor (MRCA)', wat staat voor de dichtstbijzijnde gemeenschappelijke voorouder. Het is dus de eerste persoon die twee of meer mensen delen in hun stamboom. Voor mijzelf en mijn broer zijn dit onze ouders. Voor mijn Genealogie maat Ruud, is het onze vierde grootvader en grootmoeder Petrus Mols en Elizabeth Uijdjes, die rond 1810 geboren zijn. De MRCA is vaak de focus van genetische genealogie.

De afbeeldingen die ik in mijn boek gebruik zijn uitgebracht onder Creative Commons licenties. De rechten op de afbeelding op de kaft, alsmede de auteur zijn onbekend. We weten dat deze foto rond 1880 genomen is in Tilburg. De man zittend in het midden is Petrus Franciscus Mols mijn oudvader (4^{de}), rechts van hem, zijn oudste zoon Johannes mijn betovergrootvader en rechts van hem zijn zoon Petrus Franciscus Mols,

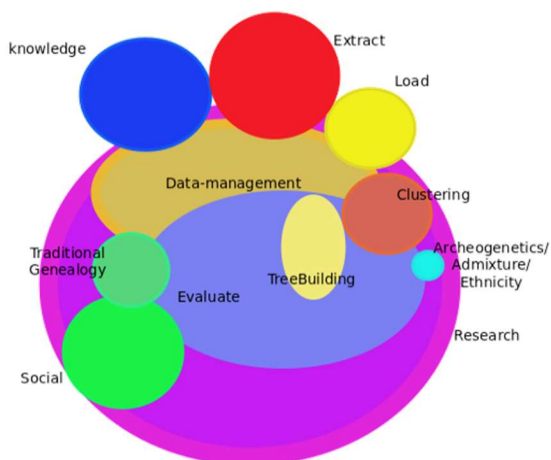
mijn overgrootvader. Allemaal mannen met wie ik DNA deel, maar die ik nooit heb mogen ontmoeten.

Ik heb ervoor gekozen om voor het boek een groot lettertype te kiezen, omdat veel van mijn lezers wat ouder zijn. Het gekozen formaat a5 om het een handzaam boek te maken, wat makkelijk gelezen kan worden. Ik heb ervoor gekozen de terminologie in het Engels te laten staan, omdat dat immers de begrippen zijn, die je op Internet tegen blijft komen. Ik besef dat papier geduldig is. Zo geduldig dat het nadeel van een papieren uitgave tot uiting kan komen, namelijk dat de inhoud verouderd. Naast dit boek heb ik een YouTube kanaal, waar ik regelmatig (Engelstalige) filmpjes post. Op het moment van schrijven staan er ruim 100 video's op. In veel van deze video's doe ik de technieken die ik in het boek uitleg voor. Het YouTube kanaal is te vinden op <https://www.youtube.com/c/erikmols>

Biezenmortel, December 2021

Het Filin Genetic Genealogy Raamwerk

Een raamwerk is een methodologie om zaken te organiseren. Net zoals je een kledingkast zou gebruiken om je garderobe te organiseren, kan je een raamwerk gebruiken om je genetisch genealogisch onderzoek te organiseren. Omdat er zoveel technieken en mogelijkheden zijn, is het belangrijk dat je een manier hebt om je onderzoek te structureren. Het raamwerk wat ik in dit boek presenteer is een manier om je onderzoek te organiseren en meer en betere resultaten te krijgen. Het boek is opgebouwd rondom dit raamwerk en alle afzonderlijke raamwerkclusters worden als hoofdstuk uitgewerkt.



Filin Genetic Genealogy Framework in overview

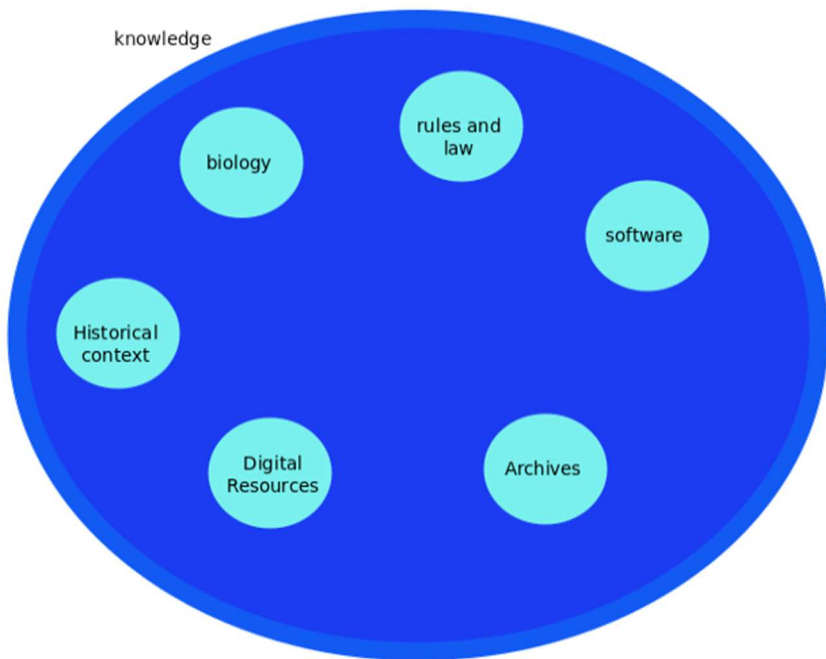
De structuur is zichtbaar in de bovenstaande afbeelding, waarbij je kan zien dat het raamwerk is opgedeeld in verschillende secties. Deze secties hebben allemaal een onderscheidende kleur en grootte. De kleur van de clusters is alleen aangebracht om het onderscheid te kunnen zien. De afmeting laat het relatieve belang van het cluster zien, ten op zichte van genetisch genealogisch onderzoek. De overlap tussen de verschillende gebieden van het raamwerk, laat zien de mate waarop deze gebieden elkaar beïnvloeden. De ‘knowledge’

(kennis) sectie gaat over theoretische kennis die je nodig hebt om een gedegen onderzoek te kunnen doen. Kennis op het gebied van biologie, computerwetenschappen en geschiedenis. De 'Extract and Load' (Extraheren en Laden) sectie gaat over het verkrijgen en verplaatsen van je DNA data, naar en tussen verschillende test bedrijven. De 'research' sectie gaat over het structureren van je onderzoek door het formuleren van een onderzoeksvraag en het kiezen van een juiste onderzoeksstrategie en methodiek, welke het beste aansluit bij jouw specifieke probleem en onderzoeksvraag. 'Data management', gaat over het opslaan van je resultaten op een dusdanige wijze zodat je ze later terug kunt vinden. Een erg belangrijke sectie is 'Clustering and Tree building' (groeperen en stamboom bouwen) wat gaat over het organiseren en structureren van je DNA matches. De 'evaluate' (evalueer) sectie combineert de verschillende lagen van het raamwerk. 'Traditional Genealogy' (klassieke genealogie), gaat over het gebruik van archieven en andere traditionele bronnen, om je

onderzoek te ondersteunen. De 'social' (sociaal) sectie gaat over het gebruiken van internet en social media, om relevante informatie over je DNA matches te vinden. De laatste sectie gaat over 'Archeogenetics, Admixture and Ethnicity' (Archeologie van genetica en afstamming / etniciteit), wat hoofdzakelijk gezien kan worden als een hulpmiddel om een ruw idee te krijgen over een geografische locatie.

Knowledge

De knowledge (kennis) sectie van het raamwerk bestaat uit verschillende disciplines.



The knowledge sectie van het Genetische Genealogie Raamwerk

Knowledge (kennis) is geen stap voor stap sectie. Het is een verzameling van relevante onderwerp gebieden die besproken worden. Het bevat uitleg en discussie over onderwerpen, die van belang zijn bij het doen van genetisch genealogisch onderzoek. Uitleg die je kunt gebruiken bij het scopen en beantwoorden van je vragen. Informatie, die je munitie en methodes geeft om jouw genetische genealogische puzzels op te lossen.

Biology (Biologie)

In deze paragraaf ga ik uitleggen wat DNA is en welke DNA mechanismen / processen belangrijk zijn voor genetische genealogie. Alle andere DNA en RNA processen laat ik buiten beschouwing, omdat deze niet van belang zijn bij het beantwoorden van genetisch genealogische vraagstukken en omdat het niet mijn bedoeling is een academisch boek over genetica te schrijven.

We hebben DNA in ons hele lichaam. DNA zit in iedere cel. Dus laten we de biologie van

genetische genealogie beginnen met het kijken naar cellen in het menselijk lichaam.

De cel, bouwsteen van het leven

Zoals je waarschijnlijk weet bestaat je lichaam uit vele organen: hersenen, hart, longen, maag, darmen en vele meer. Ook je huid en botten zijn organen. Het is zelfs zo dat je huid het grootste orgaan van allemaal is. Wanneer je naar je organen kijkt door een microscoop, dan zie je heel veel cellen. De cel is lang gezien als de primaire bouwsteen van het lichaam.

Theoretisch kun je zeggen dat, dat nu DNA is. Maar het blijft dat de cel het primaire bouwblok is, waaruit onze organen zijn opgebouwd. We hebben in ons lichaam ongeveer 37.000.000.000.000 cellen. Voor velen een getal dat je waarschijnlijk niet eens direct kan uitspreken.

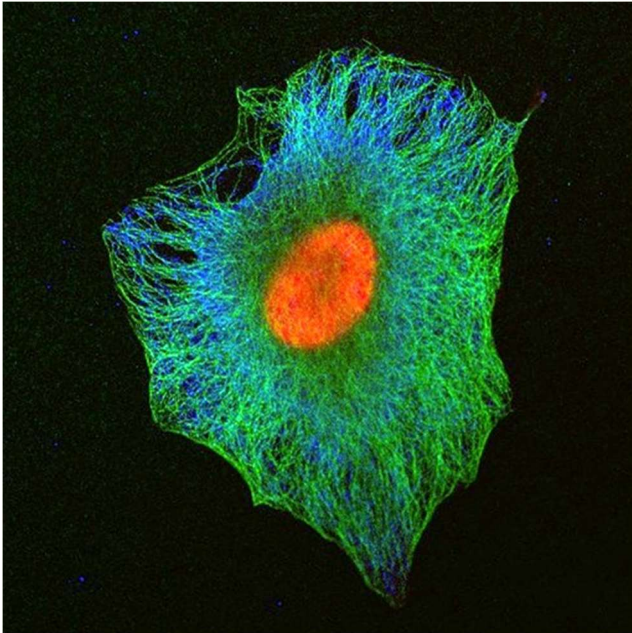
Al deze cellen bevatten een kopie van ons volledige DNA, behalve onze reproductie (seks) cellen, die bevatten slechts de helft van ons DNA. Het totaal van het DNA in een cel wordt

het genoom genoemd. Nu zou je verwachten dat wij maar één genoom hebben. Deze aanname is incorrect, in feite hebben we twee genomen. Ons autosomale / heterosomale genoom en mitochondriaal genoom, maar deze wordt later besproken.

De Cel

In de vorige paragraaf hebben gesproken over dat het menselijk lichaam uit organen bestaat en dat deze organen weer opgebouwd waren uit cellen. Wanneer je met een microscoop naar een cel kijkt, zul je wellicht wat teleurgesteld zijn. Dit omdat de verschillende onderdelen van de cel zo klein zijn, dat je deze door een normale microscoop niet kunt zien. Om deze organellen zichtbaar te maken, heb je een hele sterke microscoop nodig, een zogenaamde elektronen microscoop. Wanneer je een afbeelding van een cel, gemaakt door een elektronen microscoop bestudeerd, zie je de verschillende organellen waar een cel uit is opgebouwd. Deze organellen

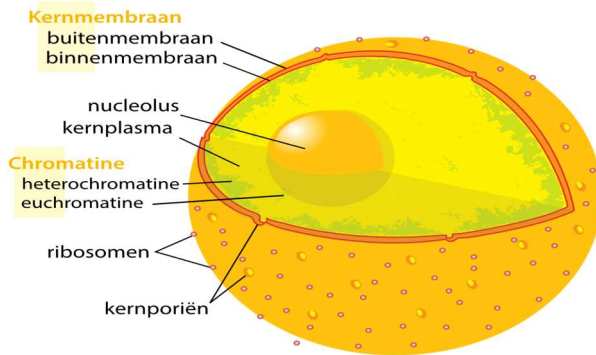
hebben één of meerdere rollen in de cel. Door de processen die de cel organellen uitvoeren kan een cel redelijk zelfstandig functioneren. Voor de genetische genealogie zijn maar twee organellen van belang: de celkern (nucleus) en de mitochondria. In de volgende afbeelding is een menselijke cel zichtbaar. De celkern is duidelijk gekleurd.



"Human cell showing nucleus and tubulin" is licensed with CC BY 4.0.

Celkern (Nucleus)

De celkern wordt meestal weergegeven in het midden van de cel. Deze nucleus is verantwoordelijk voor het controleren en sturen van de cel. Om dit te doen maakt de celkern gebruik van een soort computer programma. Dit programma wordt gelezen en de verkregen boodschappen worden als instructies rond gestuurd. Dit programma noemen we DNA. Het DNA is verspreid over 46 chromosomen. De meeste instructies zijn geschreven in RNA welke door de ribosomen vertaald wordt tot eiwitten. RNA en ribosomen zijn voor genetische genealogie verder niet interessant.



"File:Diagram human cell nucleus nl.svg" by Mariana Ruiz (LadyofHats), translated by TheBartgry is licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)

Chromosomen

Ons DNA ligt in de celkern op kleine organellen die chromosomen worden genoemd. Feitelijk zie je de bekende vorm van de chromosomen alleen maar tijdens hun opgerolde stadium, welke ze hebben tijdens de celdeling. In normale omstandigheden zijn de chromosomen uitgerold, zodat het DNA gelezen kan worden. Als een menselijk wezen hebben we 46 chromosomen. 22 paar autosomen en 1 paar heterosomen. Naast normale biologische functies zijn de heterosomen ook