

UNIVERSITATEA BABEȘ-BOLYAI
FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ

Tehnici de instruire automată aplicate în bioarheologie

Rezumatul Tezei de Doctorat

Student Doctorand: Mircea Ioan-Gabriel

Coordonator: Prof. Dr. Czibula Gabriela

2016

Cuvinte cheie: bioarheologie, hărți cu auto-organizare, arbori de decizie, învățare prin întărire, reguli de asociere relaționale, sisteme fuzzy.

Cuprins

List of publications	8
Introduction	10
1 Background	14
1.1 Multiple Sequence Alignment	14
1.1.1 Sequence to profile alignment	16
1.1.2 Needleman-Wunsch with affine gap penalty	17
1.1.3 Literature review	19
1.1.4 Protein Multiple Sequence Alignment	21
1.2 Sex identification	23
1.2.1 Literature review	24
1.3 Machine learning based models	26
1.3.1 Self organizing maps	26
1.3.2 Clustering	26
1.3.3 Decision trees	27
1.3.4 Reinforcement Learning	28
1.3.5 Relational association rule mining	29
2 A novel approach to Multiple Sequence Alignment	33
2.1 The algorithm for sequence to profile alignment	33
2.2 The proposed <i>RL</i> approach for the <i>MSA</i> problem	35
2.2.1 The learning process	36
2.3 Our approach to Protein Multiple Sequence Alignment	37
2.3.1 The reinforcement learning task	38
2.4 Computational Experiments	39
2.4.1 Experiments using the Needleman-Wunsch algorithm	39
2.4.2 Experiments using the Needleman-Wunsch with affine gap penalty algorithm	54
2.5 Experiments on PMSA using Q-learning	59
2.5.1 Artificial data sets	60
2.5.2 BALiBASE benchmarks	61

2.5.3	Discussion and Comparison to Related Work	63
2.5.4	Conclusions and future work	65
3	Novel machine learning approaches in bioarchaeology	67
3.1	Proposed approaches	67
3.1.1	An approach based on self-organizing maps	67
3.1.2	Clustering based approaches	69
3.1.3	Decision tree based approaches	70
3.1.4	A fuzzy solution to the gender detection problem	72
3.2	Experiments using self-organizing maps	78
3.2.1	First data set	78
3.2.2	Second data set	83
3.2.3	Third data set	85
3.2.4	Discussion	88
3.2.5	Analysis of the proposed SOM model	88
3.2.6	Comparison to related work	90
3.3	Experiments using clustering	93
3.3.1	Results	93
3.3.2	Discussion and comparison to related work	96
3.4	Experiments using decision trees	98
3.4.1	First case study	99
3.4.2	Second case study	99
3.4.3	Third case study	100
3.4.4	Discussion and comparison to related work	100
3.5	Experiments using fuzzy decision trees	103
3.5.1	First Case Study: South African Radius and Ulna Measurements . . .	103
3.5.2	Second Case Study: Terry Collection	103
3.5.3	Results	104
3.5.4	Discussion and comparison to related work	104
3.5.5	Conclusions and future work	105
4	Adaptive relational association rule mining	106
4.1	Methodology	107
4.2	Experimental evaluation	111
4.2.1	Experiments	111
4.2.2	Human skeletal remains data	112
4.2.3	Terry collection data set for sex identification	115
4.2.4	Synthetic data	118
4.3	Discussion	123
4.3.1	Analysis of the <i>ARARM</i> method	123

4.3.2	Comparison to related work	127
4.4	Conclusions and future work	127
5	Conclusions	129

Lista publicațiilor

Publicații în ISI Web of Knowledge

Publicații în ISI Science Citation Index Expanded

1. Gabriela Czibula, Istvan Gergely Czibula, Adela Sirbu, **Ioan-Gabriel Mircea**. A novel approach to adaptive relational association rule mining. *Applied Soft Computing journal*, Vol. 36, pp. 519-533, 2015. (**IF = 2.810**)
2. Gabriela Czibula, Vlad-Sebastian Ionescu, Diana-Lucia Miholca and **Ioan-Gabriel Mircea**. Machine learning-based approaches for predicting stature from archaeological skeletal remains using long bone lengths. *Journal of Archaeological Science*, Vol. 69, pp. 8599, 2016. (**IF = 2.196**)

Publicații în Conferințe ISI Citation Index Expanded

1. **Ioan-Gabriel Mircea** and Gabriela Czibula, and Maria-Iuliana Bocicor. *A Q-learning approach for aligning protein sequences*. Proceedings of the 11th International conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, pp. 51-58, 2015.
2. Vlad-Sebastian Ionescu and **Ioan-Gabriel Mircea** and Diana-Lucia Miholca and Gabriela Czibula. *Instance Based Learning Approaches for predicting the height of human skeletons*. Proceedings of the 11th International conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, pp. 309-316, 2015.
3. Diana Miholca, Gabriela Czibula, **Ioan-Gabriel Mircea**, and Istvan-Gergely Czibula. Machine learning approaches for gender detection of human skeletal remains. *18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, under review, 2016.
4. **I.G. Mircea** A fuzzy decision tree based method for skeletal sex determination. In *Proceedings of the Symposium on Applied Computational Intelligence, Timisoara, Romania*, pp. 447-452, 2016

5. Maria-Iuliana Bocicor, **Ioan-Gabriel Mircea**, and Gabriela Czibula. A novel reinforcement learning based approach to multiple sequence alignment. *Proceedings of the 7th International Workshop on Soft Computing Applications*, to be published, 2016.

Lucrări publicate în jurnale internaționale și prezentate la conferințe internaționale

1. **Ioan-Gabriel Mircea** and Maria-Iuliana Bocicor and Andra Dîncu. On reinforcement learning based multiple sequence alignment. *Studia Universitatis Babeș-Bolyai Series Informatica*, LIX(2):50–65, 2014. (**indexed Mathematical Reviews**)
2. **Ioan-Gabriel Mircea** and Sergiu Limboi and Mara-Renata Petrușel. A New Un-supervised Learning Based Approach for Gender Detection of Human Archaeological Remains. *Studia Universitatis Babeș-Bolyai Series Informatica*, LX(2):5–20, 2015. (**indexed Mathematical Reviews**)
3. **Ioan-Gabriel Mircea** and Gabriela Czibula and Mara-Renata Petrușel. Sex identification in archaeological remains using decision trees. *Studia Universitatis Babeș-Bolyai Series Informatica*, LX(2):91–103, 2015. (**indexed Mathematical Reviews**)
4. **Ioan-Gabriel Mircea**. An Evaluation of Basic Techniques and Methods Used in Skin Color Detection. *Studia Universitatis Babeș-Bolyai Series Informatica*, LVII(3):24–34, 2012. (**indexed Mathematical Reviews**)

Scopul cercetării întreprinse în cadrul tezei de doctorat prezentate a fost în primul rând acela de a aduce împreună două zone de cercetare aparent totale discrepante. Pe de o parte, instruirea automată este un domeniu al informaticii care încearcă în mod ambițios să construiască algoritmi capabili să învețe să rezolve probleme, folosindu-se de experiențe acumulate anterior în perioada de antrenament. Pe de altă parte, bioarheologia este la rândul ei un domeniu de cercetare foarte dinamic în care cunoștințe din domeniul biologiei, în special al biologiei moleculare dar nu numai sunt folosite în studiul populațiilor antice ale căror rămășițe pământești sunt descoperite pe șantierele arheologice. Privite de la distanță, cele două domenii de cercetare, fiecare în sine având un grad ridicat de atractivitate pentru cercetători, par însă diferite în mod fundamental.

Pe parcursul studiilor doctorale am ajuns însă la concluzia că deși într-adevăr perspectivele pe care le prezintă fiecare dintre domeniile de cercetare amintite anterior sunt cu siguranță diferite, atât domeniul instruirii automate cât și cel al bioarheologiei fiind guvernate de cutume și abordări științifice specifice, există o dependență profundă între cele două domenii de cercetare. Această conexiune va fi din ce în ce mai vizibilă pe viitor deoarece în scurt timp cele două domenii vor evolua într-o stare în care colaborarea între cercetători din ambele domenii va fi inevitabilă.

Teza de doctorat prezentată poate fi văzută ca o pledoarie pentru interdisciplinaritate. Obiectivul principal al tezei este folosirea tehnicilor de instruire automată pentru rezolvarea problemelor reale din domeniul bioarheologiei, pentru a scoate la iveală interdependențele din ce în ce mai vizibile dintre cele două domenii ce transpar în urma studiului în paralel al descoperirilor recente din ambele domenii de activitate. Este din ce în ce mai clar că, din perspectiva cercetării bioarheologice, o problemă care promite să devină din ce în ce mai profundă în viitorul apropiat este reprezentată de nevoia unor mijloace digitale de colectare, reprezentare și analiză a volumului imens de date neprelucrate din siturile arheologice care vor fi impuse de protocoalele arheologice moderne.

Un volum mare de date este însă benefic pentru cercetarea din domeniul instruirii automate deoarece cu cât algoritmii de învățare automată sunt antrenați pe un set de date mai substanțial, cu atât performanța lor va crește. Această raportare antagonică la volumul seturilor de date în fiecare dintre domeniile de cercetare reprezintă unul dintre motivele importante pentru care bioarheologia a devenit obiectul de studiu al tezei prezentate, reprezentând dovada clară că există interdisciplinaritate reală între cele două domenii de studiu atâta vreme cât ceea ce reprezintă un dezavantaj clar într-unul dintre domenii poate favoriza substanțial cercetarea din celălalt domeniu.

Alegând bioarheologia ca domeniu de interes pentru teza de doctorat, următorul pas a fost identificarea problemelor din bioarheologie care ar fi putut fi rezolvate folosind o abordare bazată pe tehnici de instruire automată. În acest demers, un ajutor semnificativ ne-a fost oferit de către Institutul de Bio Nano Științe. Ca urmare a discuțiilor purtate cu cercetătorii din institute, prima temă de cercetare abordată a fost aleasă să fie alinierea multiplă a secvențelor din biologia moleculară.

Deoarece studiile de genetică sunt folosite pe o scară din ce în ce mai largă în arheologie, și chiar mai frecvent în criminalistică, rezultatele analizelor genetice ale materialului osos obținut din siturile arheologice, de cele mai multe ori prezentat sub forma unor secvențe de nucleotide pot să scoată la iveală informații interesante atât despre individul studiat cât mai ales despre populația din care făcea parte. Un exemplu concludent care evidențiază importanța folosirii metodelor de cercetare din biologia moleculară în arheologie este cercetarea prezentată în articolul [SKW⁺10] care se bazează pe diferențele în alcătuirea moleculară a diferitelor specii de curcani din sudul Statelor Unite ale Americii pentru a dovedi că indienii americani au fost cei care îl domesticiseră înainte de venirea lui Columb.

Pentru a se putea însă obține rezultate atât de convingătoare cu o acuratețe științifică, analiza materialului genetic trebuie să fie făcută cât mai minuțios cu putință iar acest lucru devine aproape imposibil în contextul volumului mare de date care trebuie analizate. Similitățile sau diferențele dintre secvențe devin vizibile când acestea sunt aliniate astfel încât structura lor relevă în mod natural zonele în care diferă. De obicei aceste diferențe sunt rezultatul unor mutații care duc la apariția unor noi specii cu comportamente sau aspecte diferite. De aceea, alinierea multiplă a secvențelor are o importanță majoră în determinarea cu acuratețe a diferențelor din materialul genetic extras din siturile arheologice.

Am încercat astfel să găsim o rezolvare bazată pe instruire automată pentru problema alinierii mai multor secvențe de nucleotide concomitent. Cum problema este mai degrabă una legată de biologie decât de arheologie, existau deja câteva abordări în literatura de specialitate, majoritatea folosind algoritmi genetici. Este cunoscut faptul că alinierea mai multor secvențe de nucleotide este o problemă NP-completă și de aceea s-a hotărât că o abordare care ar putea avea succes ar presupune căutarea inteligentă a spațiului de căutare. Cum problema a fost formalizată ca una de găsim o permutare optimă de alinieri succesive în pereche a secvențelor, o abordare folosind învățarea prin întărire a fost preferată în detrimentul abordărilor cu algoritmi genetici. Astfel s-a concretizat prima direcție de lucru din cadrul doctoratului sub forma dezvoltării unui algoritm bazat pe învățarea prin întărire pentru rezolvarea problemei alinierii mai multor secvențe de nucleotide.

Pe de altă parte, biologia moleculară nu este singurul mod prin care biologia modelează știința modernă a arheologiei. Analiza osteologică a materialului osos dintr-un sit arheologic are o importanță majoră deoarece ajută la identificarea atributelor biologice de bază. Aceste atribute (sex, vârstă, rasă, statură, greutate) sunt de asemenea foarte importante și în criminalistică, fiind cei mai importanți indicatori folosiți la identificarea persoanelor dispărute. În criminalistică, determinarea acestor indicatori de bază este o sarcină dificilă dar în cazul bioarheologiei dificultatea crește și mai mult din cauza vârstei înaintate a scheletelor studiate, al gradului de erodare ridicat sau al incompletitudinii scheletelor. A doua direcție majoră de studiu a fost dezvoltarea unor algoritmi de instruire automată pentru stabilirea sexului sau a vârstei sau pentru aproximarea staturii.

Scopul principal al acestei abordări a fost rezolvarea sarcinilor de clasificare sau regresie amintite anterior cu o acuratețe cât mai ridicată doar pe baza măsurătorilor osteologice (

cateodată îmbogățite cu informații suplimentare de natură moleculară, istorică etc.) deoarece stabilirea acestor caracteristici de baza reprezintă punctul de temelie al cercetării bioarheologice depinzând de acuratețea acestei determinări inițiale. Majoritatea metodelor folosite în literatură chiar și la ora actuală formule matematice relativ simple sau protocoale realizate empiric și pot duce cu ușurință la erori. De aceea, pare că abordarea de rezolvare a problemei cea mai cinstită ar fi ar trebui să folosească datele osteologice în forma lor inițială și nu aproximări sau discretizări ulterioare ale acestor valori. Astfel de abordări ar trebui să poată să se adapteze la inexactitatea datelor sau la cazurile neconcludente care fac clasificarea mult mai anevoioasă.

O a treia direcție de studiu a constat în adaptarea metodelor de instruire automată la nevoile domeniului bioarheologie și aici a fost atins cu adevărat nivelul de interdisciplinaritate scontat, care a dus la implementarea unor algoritmi originali modelați pentru a rezolva sarcinile din domeniul bioarheologiei dar nu doar atât, deoarece anumiți algoritmi propuși au o importanță semnificativă și în procesul de vizualizare a datelor arheologice într-o formă cât mai intuitivă și de analiză ulterioară pe baza datelor și abstractizărilor acestora.

Teza prezintă principalele aspecte ale efortului de cercetare depus în ultimii trei ani și, pentru a evidenția modul în care fiecare dintre cele trei direcții de cercetare a fost urmărită a fost structurată după cum urmează, începând cu un capitol introductiv menit să evidențieze contextul și motivația care au alimentat întregul proces de cercetare.

Primul capitol urmează după introducere și prezintă conceptele de bază și ideile de rezolvare care vor fi folosite pe tot parcursul tezei în toate cele trei direcții de studiu. Pe de altă parte, pentru fiecare dintre direcțiile majore de studiu sunt prezentate principalele referințe bibliografice de valoare din domeniu.

Următoarele trei capitole prezintă, pe rând, cele trei abordări noi pentru rezolvarea problemelor de bioarheologie prezentate succint în secțiunile anterioare. Capitolul doi prezintă pe larg soluția bazată pe învățare prin întărire care folosește q-learning pentru a determina permutarea optimă de alinieri complete succesive în pereche astfel încât să se maximizeze scorul global de aliniere.

Următorul capitol urmărește determinarea sexului pe baza măsurătorilor neprelucrate obținute din situl arheologic folosind tehnici de învățare automată adaptate la cerințele impuse de domeniul de cercetare al bioarheologiei. Una dintre metode, cu aplicații reale în vizualizarea datelor bioarheologice, este reprezentată de hărțile cu auto-organizare folosite pentru reprezentarea într-un spațiu cu dimensionalitate redusă a datelor multi-dimensionale care caracterizează indivizii din siturile arheologice. Cum determinarea sexului poate fi privită ca o problema de grupare (clusterizare) de entități pe baza caracteristicilor comune, unele metode de clustering au fost adaptate pentru a rezolva problema. Nu în ultimul rând problema este abordată cu ajutorul unor metode bazate pe arborii de decizie. Inițial, s-a încercat o soluție care folosea arborii de decizie clasici, pentru ca mai apoi o analiză statistică atentă a datelor privind măsurătorile osteologice a scos la iveală anumite caracteristici intrinseci ale datelor arheologice care au inspirat o nouă variantă de arbore de decizie care înglobează

concepte fuzzy în construcție și obține rate de detecție foarte competitive.

Cea de-a treia direcție de studiu este prezentată în capitolul următor și detaliază folosirea regulilor de asociere în construirea unui algoritm adaptiv în rezolvarea aceleiași sarcini prezentate anterior - determinarea sexului - printr-o cu totul altă metodă decât cele prezentate înainte. Concluziile și direcțiile viitoare de studiu sunt prezentate în ultimul capitol, punându-se accent pe o direcție viitoare de studiu care se bazează pe combinarea metodelor de clasificare prezentate în ultimele două capitole în realizarea unui arbore de decizie care să se bazeze pe reguli de asociere relaționale pentru a crește acuratețea clasificării.

Dorim să evidențiem că procesul de cercetare a condus la câteva contribuții originale în domeniul bioarheologiei pe care le vom prezenta succint după cum urmează.

Prima contribuție este reprezentată de metoda de aliniere a mai multor secvențe de nucleotide bazată pe învățare prin întărire [BMC16] și o comparație cu alte abordări din literatură [BMC16, MBD14]. Totodată o analiză detaliată a efectului variației parametrilor și a folosirii algoritmului Needleman-Wunsch în varianta afină a fost efectuată în [MBD14]. Metoda propusă a fost totodată adaptată pentru lucrul cu secvențe proteice în [MCB15].

A doua contribuție originală este reprezentată de metodele de instruire automată folosite în determinarea sexului pe baza măsurătorilor osteologice *self organizing maps, clustering* and *decision trees* [CMMM16], [MLs15], [MCs15] și de evaluarea performanței lor comparativ la abordările din literatură [CMMM16, MLs15, MCs15].

A treia contribuție privește utilizarea tehnicilor de identificare a regulilor de asociere relaționale și impactul pe care aceste reguli îl au în studiul bioarheologiei și a fost realizată în articolul [CCSiM15].

În concluzie, după o muncă de cercetare de aproape trei ani, rezultatele obținute pot constitui un pas important în adaptarea algoritmilor de învățare automată la nevoile domeniului de cercetare al bioarheologiei. Nevoile informaționale din domeniu au crescut în mod drastic în ultimii ani fapt care s-a reflectat și pe parcursul cercetării curente. În urmă cu trei ani, dezavantajul major în încercarea de a găsi soluții bazate pe învățarea automată era lipsa semnificativă de seturi de date pentru antrenare. Însă, pe parcursul ultimilor ani, au apărut din ce în ce mai multe surse de date și din ce în ce mai consistente precum colecția Terry oferită de Muzeul Smithsonian sau bazele de date cu materiale genetice din ce în ce mai prezente online.

De obicei problemele propuse în domeniul bioarheologiei prezintă un grad ridicat de dificultate care sugerează o abordare bazată pe învățare automată, și dacă observăm această predispoziție în contextul creșterii susținute a măsurătorilor osteologice efectuate în arheologie și a importanței tot mai mari pe care o are antropologia în știința istoriei moderne, putem presupune că în viitorul apropiat arheologii vor fi copleșiți de volumul imens de informație pe care îl dețin și fără unelte digitale care să facă față unui volum atât de mare de date și care să fie capabile să ofere abstractizări și analize pertinente care să înlesnească munca arheologică, această zestre informațională nu va putea fi gestionată la parametri optimi.

Îndrăznim să afirmăm că viitorii cercetători din domeniul bioarheologiei vor fi nevoiți

să posede cunoștințe minime de data mining și să cunoască uneltele informaționale care să înlesnească analiza datelor, iar echipele care vor dezvolta acele unelte vor fi formate din membri care să cunoască la un nivel satisfăcător realitatea bioinformatică și anume să posede cunoștințe de genetică și biologie moleculară, antropologie, osteologie etc. pentru a putea să adapteze algoritmi necesari la cerințele impuse de domeniul bioarheologiei.

În acest context, consideram că lucrarea reprezintă un punct de început pentru viitoare cercetări interdisciplinare în domeniul bioarheologiei. Având scopul inițial de a rezolva probleme de bioarheologie cu ajutorul instruirii automate, considerăm că a reușit să dovedească apropierea reală a celor două domenii aparent total incompatibile. Munca de cercetare în acest domeniu interdisciplinar a scos la iveală atât avantajele comune pe care abordările computaționale le aduc dar și provocările pe care natura problemelor de bioarheologie le impun. Avantajele comune vor fi și mai evidente odată cu apariția uneltelor digitale care să fie folosite pe scară largă în domeniu, care va aduce cu sine o creștere semnificativă a volumului informațional de asemenea, fapt care va ajuta la obținerea de performanțe din ce în ce mai bune din partea algoritmilor bazați pe instruire automată folosiți. Interdisciplinaritatea aduce însă cu sine probleme de comunicare pe care însă considerăm că în cadrul acestor studii doctorale le-am depășit într-o oarecare măsură deoarece am stabilit relații bune de studiu atât cu cercetătorii de la Institutul de Bio Nano Științe cât și cu domnul profesor David Hunt de la Muzeul Smithsonian.

Un alt aspect important al interdisciplinarității a fost dialogul cu publicațiile de profil. Faptul că una dintre lucrările propuse în cadrul cercetării noastre a fost acceptat în *Journal of Archaeological Science* o importantă publicație din domeniu reprezintă o confirmare că o punte de comunicare a fost creată între cercetătorii din domeniul inteligenței artificiale și bioarheologi și ca un limbaj comun începe să ia ființă.

Este încurajator să vedem că domeniul de studiu al bioarheologiei modelează algoritmi de instruire automată în forme noi, inspirate din domeniu, dar care vor putea fi cu ușurință adaptate și pentru alte domenii. În crearea arborelui de decizie fuzzy, cunoștințele din domeniul bioarheologiei au fost cele care au folosit la crearea funcțiilor fuzzy atât de folosite pentru clasificări, iar această nouă formă originală va putea fi, pe viitor, să înglobeze și reguli de asociere fuzzy care vor contribui la o înțelegere mai bună a relațiilor care apar intrinsec în datele arheologice și care nu ar fi putut fi descoperite decât cu ajutorul unei abordări computaționale. Astfel, într-o simbioză perfectă, domeniul informaticii îl modelează pe cel al bioarheologiei prin uneltele computaționale pe care le oferă, iar bioarheologia modelează tehnicile computaționale folosite în forme noi, potrivite pentru rezolvarea problemelor specifice.

Bibliography

- [BMC16] Maria-Iuliana Bocicor, Ioan-Gabriel Mircea, and Gabriela Czibula. A novel reinforcement learning based approach to multiple sequence alignment. In *Proceedings of the 7th International Workshop on Soft Computing Applications*, 2016.
- [CCSiM15] G. Czibula, I.G. Czibula, A. Sîrbu, and I.G. Mircea. A novel approach to adaptive relational association rule mining. *Applied Soft Computing*, 36:519–533, 2015.
- [CMMM16] G. Czibula, I.G. Mircea, D.L. Miholca, and R. Moldovan. Machine learning approaches for gender detection of human skeletal remains. page under review, Timisoara, Romania, 2016. IEEE Computer Science.
- [MBD14] Ioan-Gabriel Mircea, Maria-Iuliana Bocicor, and Andra Dîncu. On reinforcement learning based multiple sequence alignment. *Studia Universitatis Babeș-Bolyai Series Informatica*, LIX(2):50–65, 2014.
- [MCB15] I.G. Mircea, G. Czibula, and M.I. Bocicor. A q-learning approach for aligning protein sequences. In *Proceedings of the 11th International conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania*, page to appear, 2015.
- [MCs15] Ioan-Gabriel Mircea, Gabriela Czibula, and Mara-Renata Petrușel. Sex identification in archaeological remains using decision trees. *Studia Universitatis Babeș-Bolyai Series Informatica*, LX(2):91–103, 2015.
- [MLs15] Ioan-Gabriel Mircea, Sergiu Limboi, and Mara-Renata Petrușel. A new unsupervised learning based approach for gender detection of human archaeological remains. *Studia Universitatis Babeș-Bolyai Series Informatica*, LX(2):5–20, 2015.
- [SKW⁺10] Camilla F. Speller, Brian M. Kemp, Scott D. Wyatt, Cara Monroe, William D. Lipe, Ursula M. Arndt, and Dongya Y. Yang. Ancient mitochondrial DNA analysis reveals complexity of indigenous North American turkey domestication. *Proceedings of the National Academy of Sciences*, 107(7):2807–2812, February 2010.