Proposed Study: "The Fairy Folk Hypothesis – Testing for Genetic Signatures Associated with Regions of High Fairy Intermarriage Folklore in the British Isles"

Research Question

Do populations in regions with strong traditions of fairy-human intermarriage (e.g., Myddfai in Wales, Isle of Skye, Guernsey, the Hebrides) show detectable genetic differences compared to neighboring control regions with few such traditions—differences that might reflect admixture from distinct source populations?

PHASE 1: Study Design and Site Selection (Months 1–3)

A. Define "High Fairy Intermarriage" Regions

Based on folklore sources, candidate "high tradition" regions would include:

Region	Specific Area	Folklore Basis
Wales	Myddfai / Llyn y Fan Fach	Physicians of Myddfai, fairy ancestress
Scotland	Isle of Skye (Dunvegan)	Clan MacLeod, Fairy Flag tradition
Scotland	Outer Hebrides	Multiple abduction/intermarriage tales
Guernsey	Whole island	Fairy invasion legend, "fairy bloodline"
Isle of Man	Various sites	Tynwald, Fairy Bridge traditions
Ireland	Loughcrew / Connemara	Sidhe intermarriage traditions

B. Define Control Regions

For each target region, select matched control regions with:

•Similar geography and isolation
•Comparable historical population density
•**Minimal** fairy intermarriage folklore
•Example: Mainland Argyll as control for Hebrides; lowland Wales as control for Myddfai

C. Sample Size Calculation

To detect subtle admixture signals (e.g., 5–10% contribution from a distinct source population):

Sample Type	Per Region	Total (12 regions)
Modern DNA samples	100	1,200
Ancient remains (if available)	10–20	120–240

Power analysis basis: The Viking admixture study detected 23–28% Norwegian ancestry in Shetland with ~2,500 samples . For smaller expected signals, larger samples may be needed.

PHASE 2: Sample Acquisition (Months 4–12)

A. Modern DNA Collection

Method: Saliva kits or blood samples from volunteers with:

•Grandparents born in the target region (to ensure localized ancestry)
•Informed consent and ethical approval

Costs:

Item	Per Sample	Total (1,200 samples)
Collection kit	$30	$36,000
Shipping/handling	$10	$12,000
Personnel (fieldwork)	—	$50,000
Ethics board fees	—	$5,000
Subtotal		**$103,000**

B. Ancient DNA Collection (Optional but Powerful)

If burial sites exist in target regions, sample petrous bones or teeth from individuals predating the folklore period (e.g., medieval or earlier). The **petrous bone** is ideal as it yields the highest quality ancient DNA .

Item	Per Sample	Total (150 samples)
Excavation/access	—	$30,000
Sampling/export permits	—	$10,000
Subtotal		**$40,000**

PHASE 3: Laboratory Processing (Months 13–18)

A. Wet Lab Protocol

Follow established ancient/modern DNA protocols :

1. **DNA extraction** from samples
2. **Library preparation** (double-stranded for ancient DNA)
3. **Hybridization capture** for targeted SNPs
4. **Sequencing** on Illumina platform

B. Cost Breakdown

Item	Modern (1,200)	Ancient (150)	Total
DNA extraction	$50/sample = $60,000	$100/sample = $15,000	$75,000
Library prep	$100/sample = $120,000	$150/sample = $22,500	$142,500
Capture/enrichment	$150/sample = $180,000	$200/sample = $30,000	$210,000
Sequencing (30× coverage)	$500/genome = $600,000	$1,000/genome = $150,000	$750,000
Consumables/reagents	—	—	$50,000
Personnel (lab techs)	—	—	$150,000
Subtotal			**$1,377,500**

Sequencing cost note: As of 2022, a human genome can be sequenced for approximately **$525** . Our estimate includes some cushion for library preparation and capture.

PHASE 4: Bioinformatics Analysis (Months 19–24)

A. Data Processing Pipeline

1. **Read mapping** to human reference genome (using BWA)
2. **Authentication** (for ancient DNA: damage patterns, contamination checks)
3. **Variant calling** (SNPs, indels)
4. **Quality control** (remove low-quality samples, related individuals)

B. Population Genetics Analyses

Analysis	Purpose	Software	Citation
Principal Component Analysis	Visualize population structure	EIGENSOFT	
ADMIXTURE	Estimate ancestral components	ADMIXTURE	
fineSTRUCTURE	Identify fine-scale clustering	fineSTRUCTURE	
f3-statistics	Test for admixture	ADMIXTOOLS	
qpAdm	Model admixture proportions	ADMIXTOOLS	
ROH analysis	Detect isolation/inbreeding	PLINK	
EEMS	Map gene flow barriers	EEMS	

C. Key Tests for the "Fairy Folk" Hypothesis

1. **Compare high-tradition vs. control regions** using PCA and ADMIXTURE
2. **Test for admixture** using f3-statistics of form f3(Target; Control1, Control2)
3. **If admixture detected**, model proportions with qpAdm
4. **Search for unique haplotypes** enriched in high-tradition regions
5. **Compare modern to ancient DNA** from same regions to detect population continuity or replacement

D. Cost Breakdown

Item	Cost
Computing infrastructure (cloud/cluster)	$50,000
Software licenses (if any)	$5,000
Bioinformatics personnel (2 years)	$200,000
Subtotal	**$255,000**

PHASE 5: Interpretation and Publication (Months 25–30)

A. Expected Outcomes

Possible Result	Interpretation
No genetic difference	Fairy traditions are purely cultural/supernatural; no biological basis
Subtle structure	Fairy regions show isolation-by-distance or founder effects, not admixture
Detectable admixture	Signal from a distinct source population—possibly pre-Celtic, Norse, or other
Unique haplotypes	Could indicate ancient population remnants (e.g., Pictish?)

B. Cost Breakdown

Item	Cost
Manuscript preparation	$10,000
Publication fees (open access)	$3,000
Conference presentations	$10,000
Subtotal	**$23,000**

TOTAL COST AND TIME ESTIMATE

Phase	Duration	Cost
Phase 1: Design	3 months	$0 (in-kind)
Phase 2: Sample acquisition	9 months	$143,000

Phase	Duration	Cost
Phase 3: Laboratory	6 months	$1,377,500
Phase 4: Bioinformatics	6 months	$255,000
Phase 5: Publication	6 months	$23,000
Contingency (15%)	—	**$269,625**
TOTAL	**30 months (2.5 years)**	**~$2,068,125**

Alternative: Phased Approach (Lower Cost)

If full funding is not available, a scaled-down version could be:

Phase	Description	Cost
Pilot study	2 target regions + 2 controls; 200 modern samples only	~$350,000
If pilot shows signal	Expand to all regions	~$800,000 additional
Add ancient DNA	If sites available and funding permits	+$400,000

Feasibility Assessment

Strengths of This Design

•**Proven methodology**: Similar approaches detected Viking admixture in Shetland (23–28%)
•**Fine-scale resolution**: Modern methods can detect structure across as little as 3 km
•**Ancient DNA integration**: Petrous bone methods yield high-quality data even from poor samples

Challenges

•**Defining "fairy regions"** is subjective; requires careful folklorist collaboration
•**Signal may be extremely subtle** or nonexistent
•**Ancient samples** may be unavailable in target regions
•**Ethical considerations**: Must engage with local communities respectfully

---

Key References Supporting This Design

1. **Sampling methods**: Petrous bone optimal for aDNA
2. **Sequencing costs**: ~$525 per genome as of 2022
3. **Population structure**: Proven success detecting Viking admixture
4. **Analytical tools**: qpAdm and f3-statistics validated for admixture detection
5. **Protocols**: Comprehensive pipeline available