Phylogenetic assessment of the evolution of the fictional races of Tolkien across multiple fictional universes

Jermaine Mahguib

EEOB 563, ISU Spring 2019

INTRODUCTION

The high fantasy races of J. R. R. Tolkien, as he described them, became the basis for races that inhabit numerous universes in modern fiction. Here the term ‘race’ is used freely to refer to fictional, reproductively independent species of intelligent human-like peoples inhabiting fantastical worlds; the term is used ubiquitously in fantasy and science-fantasy based literature, and commonly employed in open discussions on these intellectual properties (IPs).

Permutations of Tolkien’s elves, dwarves, humans and orcs are found across a broad range of written literature and published video game series, each adding new physical features, behavioral ranges and story roles, while retaining others from the original descriptions. For example, in Tolkien’s fictional works, the elves fill the archetypal role of the ‘old man,’ the ancient first race to inhabit the world. In the contemporary timelines of Tolkien, the elves are a race whose time of prominence has come and gone, and they serve as a fount of wisdom on matters of history and ancient arcana, and are often thrust into the role of providing some form of guidance for younger races. This archetypal role is often maintained in derivatives of the race across fictional universes.

Other examples of retained traits include the dwarves’ affinity for occupying subterranean domains, the humans’ tendency to expand their empires, and the orcs’ green skin and ‘monstrous’ visages. Conversely, a number of traits have seen some drift from Tolkien, such as the role that the orc race plays in stories. While Tolkien’s orcs were portrayed as vile and heinous, and this persisted across other IPs, some universes established their orc peoples as being more on par with their other ‘noble’ races (e.g. humans, elves and dwarves).

In the present study, a suite of traits that describe the core races of several fictional universes was established and a character state matrix was created to provide a dataset for phylogenetic analysis in order to assess how these races have evolved from Tolkien’s original incarnations. Character traits were categorized into three classifications: physiological traits, behavioral traits, and archetypal traits. For example, comparative physical *strength* of typical members of a given group was categorized as a physiological trait, while *aggression* level was categorized under behavioral traits.

Character state matrix data was used in a set of maximum likelihood analyses in an attempt to address two primary questions: (1) do various incarnations of Tolkien’s fictional races from different IPs form statistically supported phylogenetic clades; (2) in what categorical respects (physiological, behavioral or archetypal) have the original high-fantasy races of Tolkien diverged the most across multiple incarnations?

The former was approached via a maximum likelihood analysis using character state sequence data from all three trait categories concatenated. If the tested IPs exhibit minimal alteration to categorical character traits describing their fictional races relative to the type races of Tolkien, then we can expect to see, for example, all elven derivatives forming a supported clade in the tree topology. If, however, the fictional peoples of fantasy IPs have substantially deviated in our established character traits across universes, then we can expect the topology of the tree to suggest few if any supported clades, and instead observe a tree characterized by polytomies.

To address the second question, phylogenetic trees were generated using categorical sequence alignments separately, and compared to one another; in other words, trees were estimated using aligned character state sequences from the physiological traits category, others from the behavioral traits and archetypal traits categories. If a category of character traits has experienced less deviation from the original Tolkien descriptions, then we would expect to see distinct clades in the tree topology, were all elves form a group, all dwarves, etc. If, on the other hand, some incarnations of Tolkien’s fictional races have experienced substantial categorical deviation, then we might expect to see more polytomies or a single large polytomy in the tree topology.

METHODS

*Taxa and Multistate Sequence Data*

A total of 70 traits (32 physiological, 20 behavioral and 18 archetypal) were considered for this study, each assigned a set of numerically coded states (see Tables 1 – 3). The number of different states per trait varied from 2 to as many as 7, though the majority of traits were coded as having between 3 to 5 states. The 70 character traits were scored for a total of 32 taxa hailing from one of seven fictional universes: 4 taxa from J. R. R. Tolkien’s *The Lord of the Rings*; 5 from Ed Greenwood’s Dungeons & Dragons-inspired *The Forgotten Realms* series of novels and games, produced by various authors and publishers; 5 from Bethesda Studios’ *The Elder Scrolls* series; 5 from Blizzard Entertainment’s *Warcraft* series; 5 taxa each from Games Workshop’s *Warhammer* and *Warhammer 40K* series of novels and games, produced by varies authors and publishers; and 3 taxa from CD Projekt Red’s *The Witcher* series, adapted from the novels by Andrzej Sapkowski.

Four additional taxa were selected and scored for use as outgroups in phylogenetic inference: 2 taxa from Blizzard Entertainment’s *Starcraft* series; an additional taxon from Games Workshop’s *Warhammer 40K* series; and 1 taxon from film series, novels and graphic novels based on Dan O’Bannon and Ronald Shusett’s *Alien*. All taxa, the respective fictional universes to which they belong, and the numerical sequence of trait scores extracted from the character state matrix, are shown in Table 4. All character states in a given category, for a given race, were determined relative to other members of the same universe; cross-universe comparisons were not considered when determining character states for taxa.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| state matrix. The scores of the matrix were used as sequence alignments for phylogenetic inference. |  |  |  | 7 = variable |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 6 = varies among members of group |  | 6 = black |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 5 = 1 ft or more taller than average human | 5 = varies among members of group | 5 = golden skin tones |  |  |  |  |  |  |  |  |  |  |  |  |  | 5 = no ears |  |  |  |  |  |  |  |  |  |  |  |  | 5 = varies among members of group | 5 = low to high magical aptitude | 5 = low to high psychic aptitude |
|  | 4 = slightly taller than average human | 4 = bulky and muscular | 4 = purple or blue or grey skin tones |  |  |  |  |  |  | 4 = varies among members of group |  |  |  |  |  |  | 4 = very elongated |  |  |  |  |  | 4 = monstrous | 4 = very high strength | 4 = very high agility | 4 = very high intelligence | 4 = very high mental fortitude | 4 = very high hardiness | 4 = very high perception | 4 = functionally immortal | 4 = moderate to high magical aptitude | 4 = moderate to high psychic aptitude |
|  | 3 = average human | 3 = stocky and muscular | 3 = green or grey skin tones |  |  | 3 = varies among members of group | 3 = varies among members of group | 3 = varies among members of group | 3 = varies among members of group | 3 = rolling | 3 = no eyes |  | 3 = no teeth |  |  | 3 = no ears | 3 = elongated |  | 3 = characteristic |  |  | 3 = sexually monomorphic | 3 = highly attractive | 3 = high strength | 3 = high agility | 3 = high intelligence | 3 = high mental fortitude | 3 = high hardiness | 3 = high perception | 3 = very long-lived (thousands of years) | 3 = moderate magical aptitude | 3 = moderate psychic aptitude |
| Table 1. List of physiological character traits and numerically-coded associated states used to score the character- |  | 2 = slightly shorter than average human | 2 = lean and muscular | 2 = full range of human skin tones | 2 = exoskeleton | 2 = elongated | 2 = terminate in claws/scythes | 2 = 4 digits per limb | 2 = no | 2 = digitigrade | 2 = limbless | 2 = glowing, energy-infused eyes | 2 = no | 2 = sharpened teeth | 2 = no | 2 = no | 2 = pointy | 2 = slightly elongated | 2 = no | 2 = minimal | 2 = no | 2 = asexual | 2 = sexually dimorphic | 2 = moderate to highly attractive | 2 = moderate to high strength | 2 = moderate to high agility | 2 = moderate to high intelligence | 2 = moderate to high mental fortitude | 2 = moderate to high hardiness | 2 = moderate to high perception | 2 = long-lived (hundreds of years) | 2 = low to moderate magical aptitude | 2 = low to moderate psychic aptitude |
| **Scores** | 1 = 1 ft or more shorter than average human | 1 = lean and lithe | 1 = human fair skin tones | 1 = endoskeleton | 1 = globose human-like | 1 = terminate in digits | 1 = 5 digits per limb | 1 = yes | 1 = plantigrade | 1 = legged | 1 = physical, human-like eyes | 1 = yes | 1 = flat teeth | 1 = yes | 1 = yes | 1 = lobed like humans | 1 = human-lengthed | 1 = yes | 1 = none | 1 = yes | 1 = sexual | 1 = androgynous | 1 = unattractive | 1 = low to moderate strength | 1 = low to moderate agility | 1 = low to moderate intelligence | 1 = low to moderate mental fortitude | 1 = low to moderate hardiness | 1 = low to moderate perception | 1 = short-lived (human lifespan or less) | 1 = none | 1 = none |
| **Physiological Traits** | Height | Body Type | Skin Tone | Skeletal Structure | Cranium | Type of Forelimbs | Number of Forelimb Digits | Opposable Thumbs Present | Type of Legs | Type of Locamotion | Type of Eyes | Mouth Present | Type of Teeth | Tusks Present | Ears Present | Ear Shape | Ear Length | Hair Growth on Head | Facial/Body Hair | Nose Present | Reproduction | Sexual Appearance | Attractiveness | Strength | Agility | Intelligence | Mental Fortitude | Hardiness | Perception | Longevity | Magical Aptitude | Psychic Aptitude |

*Trait Categorical Descriptions*

The 70 character traits used in this study were divided up into three categories, *physiological* traits, *behavioral* traits and *archetypal* traits, each of which warrants some additional elaboration here. The physiological category consists of traits that characterize physical descriptions of typical members of a given fictional race/taxon. Examples include physical characteristics such as the typical height of taxon members, the type of physical locomotion they utilize to move around in 3D space, the presence and type of, or absence, of physical features such as ears and hair, traits associated with physical or mental performance such as strength and intelligence of typical taxon members, as well as traits related to fictionally established physiology such as longevity and aptitude for supernatural expression.

The behavioral category consists of traits that characterize some generalized behaviors associated each fictional race as a whole. Some examples include behavioral adaptability of typical taxon members, referring to the ability of individuals to adapt or adjust their general thinking and/or behavior to different circumstances, a race’s typical attitude toward matters of foreign policy, the general

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| associated states used to score character-state matrix. Scores of matrix were used as sequence alignments for phylogenetic inference. |  |  |  |  |  |  | 5 = feral |  |  |  | 5 = invasive |  |  |  |  | 5 = very high |  | 5 = not expressed | 5 = not expressed | 5 = not expressed |  |
|  | 4 = highly adaptable | 4 = extremely adaptable | 4 = tolerant | 4 = highly aggressive |  | 4 = exquisite |  | 4 = severe exploitation | 4 = highest grade | 4 = war and conquest | 4 = prescient |  | 4 = hive mind | 4 = display all-consuming spirituality | 4 = moderate to high |  | 4 = excessively so | 4 = excessively so | 4 = excessive | 4 = none expressed |
|  | 3 = adaptable | 3 = highly adaptable | 3 = moderate or grudging tolerance | 3 = moderate to high aggression | 3 = easily provoked into aggression | 3 = high to very high | 3 = high engagement | 3 = high exploitation | 3 = high grade | 3 = trade, commerce and conquest | 3 = moderate to high levels | 3 = moderate to high impulsivity | 3 = socially interconnected | 3 = display broad spirituality | 3 = low to moderate | 3 = form no familial bonds | 3 = overly so | 3 = overly so | 3 = overt | 3 = overtly expressed |
|  | 2 = somewhat adaptable | 2 = adaptable | 2 = low tolerance | 2 = low to moderate aggression | 2 = provoked into aggression on threat | 2 = moderate to high | 2 = moderate engagement | 2 = moderate exploitation | 2 = moderate grade | 2 = trade and commerce | 2 = low to moderate levels | 2 = low to moderate impulsivity | 2 = lean toward social dependence | 2 = display some spirituality | 2 = low | 2 = form strong familial bonds | 2 = moderately so | 2 = moderately so | 2 = moderate | 2 = openly expressed |
| Table 2. List of behavioral character traits and numerically-coded | **Scores** | 1 = rigid | 1 = somewhat adaptable | 1 = near zero tolerance | 1 = low aggression | 1 = not easily provoked into aggression | 1 = basic to moderate | 1 = low ingagement | 1 = low exploitation | 1 = low grade | 1 = isolationists | 1 = none | 1 = low impulsivity | 1 = lean toward individual independence | 1 = none | 1 = none | 1 = form loose familial bonds | 1 = minimally | 1 = minimally | 1 = minimal | 1 = subtly expressed |
| **Behavioral Traits** | Behavioral Adaptability | Environmental Adaptability | Tolerance of Other Races | Aggression | Incitable Aggression | Martial Prowess | Explorative | Environmentally Exploitative | Craftsmenship | Foreign Policy | Prescience | Impulsivity | Sociality | Spirituality | Scholarship | Familial Bonds | Hedonistic | Sadistic | Curiosity | Joviality |

level of expression of individualistic behaviors such as joviality, impulsivity and sadism, general level of engagement in sociality and the formation of familial bonds, and engagement in practical behaviors such as martial training, crafts and predicting future events.

The archetypal category encompasses character traits and states that describe how a race fits into a particular universe in terms of the stories that are told within them. For example, in Tolkien’s stories the orcs fill the role of an irredeemably evil race of vile killers who serve as a persistent villainous army that threatens the noble races of the world. In another universe, the Warcraft universe, the orcs are initially presented as a race occupying a similarly villainous story role, but are later revealed to actually be a noble race themselves who are only pitted against the protagonist races circumstantially. These adjustments to the original archetypes are part of the evolution of these races across fiction, and thus an attempt has been made here to include these considerations in the proposed analysis. Some examples of archetypal traits featured in the present study include racial age, or how long the group has inhabited the fictional world, contemporary demographic prevalence, role in storytelling, and racial origin/creation.

The states and scoring for these traits, across all three categories, was accomplished with the understanding that individuals of any given fictional group can and do vary dramatically from what their collective race may have been scored as here for a given trait; scoring for these traits was meant to reflect broad generalities that are often used to describe these races in either their source material or in supplemental secondary material found on the internet.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| matrix. The scores of the matrix were used as sequence alignments for phylogenetic inference. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 6 = biological |  |  |
|  | 5 = "force of nature" |  | 5 = rare or extinct |  |  | 5 = hyper advanced tech |  |  | 5 = background lore |  |  |  |  |  |  | 5 = hyper advanced |  |  |
|  | 4 = evil |  | 4 = fading | 4 = influential to security | 4 = deep connection to nature | 4 = highly advanced teched | 4 = hyper evolved or created | 4 = highly susceptible | 4 = secondary focus (villainous) | 4 = often associated with | 4 = martial (biological) | 4 = unknown |  | 4 = all-consuming | 4 = seek to defeat and assimilate enemies | 4 = highly advanced | 4 = sweepingly grand | 4 = unassociated with particular habitat |
|  | 3 = antagonistic | 3 = young; "upstart" | 3 = uncommon | 3 = highly influential | 3 = high connection to nature | 3 = moderate to high tech | 3 = highly evolved or created | 3 = susceptible | 3 = primary focus (villainous) | 3 = often entangled | 3 = martial (technological & supernatural) | 3 = created by another race via tech or magic | 3 = seek to consume their environment | 3 = excessively motivated | 3 = seek to defeat and subjugate enemies | 3 = advanced | 3 = grand | 3 = underground or within mountains |
| Table 3. List of archetypal character traits and numerically-coded associated states used to score the character-state |  | 2 = pragmatic | 2 = middle aged; "long established" | 2 = moderately common | 2 = in some respects | 2 = some connection to nature | 2 = low to moderate tech | 2 = moderately evolved or created | 2 = somehwat susceptible | 2 = secondary focus (protagonists) | 2 = sometimes entangled | 2 = martial (supernatural) | 2 = product of evolution | 2 = seek to control their environment | 2 = moderate to highly motivated | 2 = seek to defeat and exterminate enemies | 2 = moderate | 2 = expansive | 2 = in forests and woods |
| **Scores** | 1 = noble | 1 = ancient; "first born" | 1 = common | 1 = minimally | 1 = none | 1 = low tech | 1 = primitively evolved or created | 1 = insusceptible | 1 = primary focus (protagonists) | 1 = rarely entangled | 1 = martial (technological) | 1 = divinely created or manipulated | 1 = seek harmony with their environment | 1 = minimally motivated | 1 = seek to defaet and hault enemies | 1 = basic | 1 = minimalist | 1 = in the open world |
| **Archetypal Traits** | Conflict Role | Racial Age | Prevalence | Culturally Influential | Connection to "Nature" | Technological Advancement | Biological Advancement | Corruptability | Story Role | Prophecy Entanglement | Threat Type | Creation | Broad Motivation | Economic Motivation | Domination Motivation | Architectural Design | Architectural Scope | Associated Homeland Habitat |

*Maximum Likelihood Analyses*

A set of maximum likelihood analyses were conducted using the program *RaxML* in order to approach addressing two questions; how much have iterations of Tolkien’s fantasy races changed across different IPs, and have they changed, if at all, more or less so with respect to one of the trait categories over the others? To get at the first question, character state sequences extracted from the trait scoring matrix from all three trait categories were concatenated for each taxa, and all 36 concatenated sequences were aligned for use in two analyses conducted in *RaxML*: (1) search of the tree space of possible topologies for the data and identification of the best tree; (2) calculation of bootstrap support values for tree nodes. Option details for the commands used in *RaxML* to run the former are as follows.

A total of 400 likelihood trees were searched (--tree pars{200},rand{200}) using a multistate model with an 8 state ceiling and equal rates of change for all possible state changes (--model MULTI8\_MK). The ‘MULTI8’ model was necessary to run this alignment because one of the physiological traits (skin tone; see Table 1) was coded with 7 possible states; in order for *RaxML* to run a multistate analysis, the model must be set to a maximum number of states that is greater than the total number of unique states that appear in the dataset by at least one. The ‘\_MK’ designates the model as having equal rates of change for all possible state changes. This was used because the assumption was made that in reality creators are essentially free to make whatever changes they want to Tolkien’s archetypes for use in their own intellectual properties; nothing technically restricts them from doing so. Three taxa from the alignment were designated as outgroups (--outgroup Zerg\_OG,Tyranids\_OG,Xenomorphs\_OG) to root trees found in the search. Originally the ‘Protoss\_OG’ taxon was intended to be used as an outgroup in addition to the other three, but *RaxML* was unable to root trees in the search when that sequence was included as an outgroup.

Tree node support values were calculated by a non-parametric bootstrap analysis performed with 1000 replicates (--bs-trees 1000), using the same outgroup assignments and multistate model as in the tree search analysis above. Convergence of bootstraps was checked post-analysis using the --bsconverge command in *RaxML*. The default cut-off value for MRE-based bootstopping tests performed by the program during convergence testing was used.

To attempt to gain an idea of how the taxa used in this study have changed comparatively between the three trait categories (physiological, behavioral and archetypal), matrix score sequences for traits from each category were aligned separately and analyzed in a similar manner as described for the concatenated alignment above, except that the multistate model utilized was adjusted appropriately depending on the alignment (while the physiological traits alignment featured a total of 7 state variables, the behavioral and archetypal alignments contained a total of 5 and 6 state variables, respectively). Additionally, the outgroups that were included were adjusted as well. *RaxML* was unable to root trees during tree searches for any of the single category alignments when any combination of the four outgroup taxa were designated as such. The program was only able to root trees searched when a single taxon was designated as the outgroup for each alignment. As a result, maximum likelihood analyses were conducted on two sets of single-category alignments, each set with a different designated single taxon outgroup.

Of the four fictional races chosen for this study to serve as outgroups, the Zerg, Tyranids and Xenomorphs taxa are all very similar to one another across a majority of the character traits scored. Therefore, one of these taxa (Zerg) was chosen to be the sole outgroup for the first set of single-category alignments (Tyranids\_OG, Xenomorphs\_OG and Protoss\_OG sequences were removed from these alignments entirely). For the second set of single-category alignments, the Protoss taxon was designated as the outgroup and sequences from the other three taxa were removed from the alignments. This replication was done in part because it was difficult to decide which of the four taxa would work best as the single outgroup for the rest of the dataset, and in part simply to see how using two different outgroups would affect the resulting tree topologies. Together with the concatenated alignment run, a total of 7 maximum likelihood analyses were conducted on datasets for this study.

RESULTS