

# Introduction

1 The publication of Jacob and Wilhelm Grimm's Children's and Household  
Tales (1812-1814) [1] two hundred years ago sparked enormous public  
and academic interest in traditional stories told among "the common  
people", and helped establish folklore as a field for serious  
5 academic inquiry. Inspired by the Grimms' methods, a new generation  
of researchers ventured outside the library and into the villages  
and households of the rural peasantry to collect colourful tales  
of magical beasts, wicked stepmothers, enchanted objects, and  
indefatigable heroes [2]. One of the most unexpected and exciting  
10 discoveries to emerge from these studies was the recurrence of many  
of the same plots in the oral traditions associated with different  
- and often widely separated - societies and ethnic groups. Thus,  
the Brothers Grimm noted that many of the ostensibly "German"  
folktales which they compiled are recognisably related to stories  
15 recorded in Slavonic, Indian, Persian and Arabic oral traditions  
[3]. These similarities have attracted the attention of folklorists,  
literary scholars, anthropologists, cognitive scientists and others  
for a variety of reasons: For example, cognate tales in other  
cultures have been studied to try and reconstruct the origins and  
20 forms of classic western fairy tales before they were first written

down [2] [4]. Other researchers have examined the distributions of common plot elements within and across regions to make inferences about past migration, cross-cultural contact, and the impact of geographical distance and language barriers on cultural diffusion [5] [6]. Last, it has been suggested that patterns of stability and change in stories can furnish rich insights into universal and variable aspects of the human experience, and reveal how psychological, social and ecological processes interact with one another to shape cultural continuity and diversity [7] [8] [9].

25 Unfortunately, since folktales are mainly transmitted via oral rather than written means, reconstructing their history and development across cultures has proven to be a complex challenge. To date, the most ambitious and sustained effort in this area has been carried out by folklorists associated with the so-called

35 "historic-geographic" school, which was established toward the end of the nineteenth century [10]. These researchers have sought to classify similar folktales from different oral literatures into distinct "international types" based on consistencies in their themes, plots and characters. The most comprehensive and up-

40 to-date reference work in this field, the Aarne-Uther-Thompson (ATU) index, identifies more than two thousand international types distributed across three hundred cultures worldwide [11]. Exponents of the historic-geographic school believed that each

international type could be traced back to an original "archetype"  
45 tale that was inherited from a common ancestral population, or  
spread across societies through trade, migration and conquest.  
Over time, the tales' original forms were then adapted to suit  
different cultural norms and preferences, giving rise to locally  
distinct "ecotypes" [5]. The historic-geographic method sought  
50 to reconstruct this process by assembling all the known variants  
of the international type and sorting them by region and chronology.  
Rare or highly localised forms were considered to be of likely  
recent origin, whereas widespread forms were believed to be probably  
ancient, particularly when they were consistent with the earliest  
55 recorded versions of the tale [2] [4].

The historic-geographic method has been criticised for a number  
of reasons [4]. First, it has been suggested that the criteria  
on which international types are based are arbitrary and  
ethnocentric. Most types are defined by the presence of just one  
60 or two plot features ("motifs"), and gloss over dissimilarities  
among tales within the same group as well as their resemblances  
to tales belonging to other groups [12]. Since the majority  
of international types were originally defined in relation  
to the western corpus, tales from other regions are often difficult  
65 to classify according to the ATU index because they lack one or  
more of the key diagnostic motifs, or fall between supposedly

distinct international types [12] [13]. A second, related problem with the historic-geographic method is sampling bias. Given that European folklore traditions have been studied far more  
70 intensively than any others, reconstructions based on the frequency and chronologies of variants are likely to be heavily skewed. Last of all, some researchers have suggested orally transmitted tales are too fluid and unstable to be classified into distinct groups based on common descent, and that the classification  
75 of international types is often superficial [13] [14]. According to this view, the aims of the historic-geographic school are at best unrealistic, if not entirely misconceived.

This study proposes a novel approach to studying cross-cultural relationships among folktales that employs powerful, quantitative  
80 methods of phylogenetic analysis. Phylogenetics was originally developed to investigate the evolutionary relationships among biological species, and has become increasingly popular in studies of cultural phenomena (dubbed "phylomemetics" [15]), including languages [16] [17] [18] [19] [20] [20], manuscript traditions  
85 [21] [22] [23] and material culture assemblages [24] [25] [26] [27] [28] [29] [30] [25] [31]. In each case, the aim of a phylogenetic analysis is to construct a tree or graph that represents relationships of common ancestry inferred from shared inherited traits (homologies). Folktales represent an excellent target

90 for phylogenetic analysis because they are, almost by definition,  
products of descent with modification: Rather than being composed  
by a single author, a folktale typically evolves gradually over time,  
with new parts of the story added and others lost as it gets passed  
down from generation to generation. Recent case studies of the urban  
95 legend 'Bloody Mary' [32], the 'Pygmalion' family of myths in Africa  
[33], and western European variants of the folktale 'The Kind and  
the Unkind Girls' [6] have demonstrated the utility of phylogenetic  
techniques for reconstructing relationships among variants within  
a given tale type. The present study aims to establish whether these  
100 methods can also be used to differentiate the tale types themselves,  
and test the empirical validity of the international type system.  
In addressing this question, phylogenetics has several advantages  
over traditional historic-geographic methods. First, rather than  
basing the classification of related tales on just a few privileged  
105 motifs, phylogenetic analysis can take into account all the features  
that a researcher believes might be relevant. Second, phylogenetic  
reconstruction does not assume a-priori that the most common form  
of a trait, or the form exhibited by the oldest recorded variant, is  
necessarily ancestral. It is therefore likely to be less vulnerable  
110 to the strong European bias in the folktale record than traditional  
historic-geographic methods. Third, phylogenetics provides useful  
tools for quantifying the relative roles of descent versus

other processes, such as convergence and contamination, in generating similarities among taxa. These include statistical techniques  
115 for measuring how well patterns in a dataset fit a tree-like model of descent [34], and network-based phylogenetic methods that have been designed to capture conflicting relationships [35], [36]. Such methods make it possible to evaluate the coherence and degree of overlap between international types indicated by the analyses.

120 The study focuses on one of the most famous and controversial international types in the folktale literature, ATU 333 - 'Little Red Riding Hood' [37] [38]. Most versions of the story in modern popular culture are derived from the classic literary tale published by Charles Perrault in seventeenth century France [39], which recounts  
125 the misadventures of a young girl who visits her grandmother's house, where she is eaten by a wolf disguised as the old woman. It is widely believed that Perrault based his text on an old folktale known simply as 'The Story of Grandmother', versions of which have survived in the oral traditions of rural France, Austria and northern Italy  
130 [38]. In many of these tales, the girl lacks her characteristic red hood and nickname, and manages to outwit the wolf before he can eat her: After finally seeing through the villain's disguise, the girl asks to go outside to the toilet. The wolf reluctantly agrees, but ties a rope to her ankle to prevent her from escaping. When  
135 she gets out, the girl cuts the rope, ties the end to a tree, and

flees into the woods before the villain realises his mistake. Another variant of the plot has the young girl - commonly named Catterinella - taking a basket of cakes to her aunt/uncle, who turns out to be a witch or werewolf. On the way there, she eats the cakes and replaces

140 them with donkey dung. When the aunt/uncle discovers her deception, (s) he comes to her house at night and devours her in bed. Although these tales were recorded long after Perrault published his version, a rediscovered 11<sup>th</sup> century poem written in Latin by a priest in Liège provides intriguing evidence that a story similar to Little Red Riding

145 Hood was circulating in parts of western Europe in medieval times [40]. The poem, which purports to be based on a local folktale, tells of a girl who wanders into the woods wearing a red baptism tunic given to her by her godfather. She encounters a wolf, who takes her back to its lair, but the girl manages to escape by taming the wolf's cubs.

150 Highly similar stories to Little Red Riding Hood have been recorded in various non-western oral literatures. These include a folktale that is popular in Japan, China, Korea and other parts of East Asia known as 'The Tiger Grandmother' [41] [42], in which a group of siblings spend the night in bed with a tiger or monster

155 who poses as their grandmother. When the children hear the sound of their youngest sibling being eaten, they trick the villain into letting them outside to go to the toilet, where, like the heroine of The Story of Grandmother, they manage to escape.

Another tale, found in central and southern Africa [43] [44], tells  
160 of a girl who is attacked by an ogre after he imitates the voice  
of her brother. In some cases, the victim is cut out of the ogre's  
belly alive - an ending that echoes some variants of Little Red  
Riding Hood recorded in Europe, including a famous text published  
by the Brothers Grimm in nineteenth century Germany [1].

165 Despite these similarities, it is not clear whether these tales can  
in fact be classified as ATU 333. Some writers [44] [45] [46] suggest  
they may belong to another international tale type, ATU 123, The Wolf  
and the Kids, which is popular throughout Europe and the Middle  
East. In this tale, a nanny goat warns her kids not to open  
170 the door while she is out in the fields, but is overheard by a wolf.  
When she leaves, the wolf impersonates her and tricks the kids  
into letting him in, whereupon he devours them. Versions of the tale  
occur in collections of Aesop's fables, in which the goat kid avoids  
being eaten by heeding the mother's instruction not to open the door,  
175 or seeks further proof of the wolf's identity before turning him  
away. In an Indian cognate of The Wolf and the Kids, known as 'The  
Sparrow and the Crow', the villain tricks the mother into letting  
her into the house, and eats her hatchlings during the night.

Although ATU 123 is believed to be closely related to ATU 333, it is  
180 classified as a separate international tale type on the basis of two  
distinguishing features. First, ATU 333 features a single victim

who is a human girl, whereas ATU 123 features multiple victims (a group of siblings) who are animals. Second, in ATU 333 the victim is attacked in her grandmother's house, while in ATU 123 the victims  
185 are attacked in their own home. However, the application of these criteria to non-western oral traditions is highly problematic: Thus, in most of the African tales the victim is a human girl (grouping them with ATU 333), but she is attacked in her own home rather than a relative's (grouping them with ATU 123). The East  
190 Asian tales also feature human protagonists (ATU 333), but they are usually a group of siblings rather than a single child (ATU 123). In most variants of the tale, they are attacked after being left at home by their mother (ATU 123), but in some cases they encounter the villain en route to their grandmother's house (as per ATU 333).  
195 The ambiguities surrounding the classification of the East Asian and African tales exemplify the problems of current folklore taxonomy. While ATU 333 and ATU 123 are easy to discriminate between in a western context, tales from other regions share characteristics with both types and do not comfortably fit the definitions of either.  
200 With that in mind, the present study addresses two key questions: Can the tales described above be divided into phylogenetically distinct international types? If so, should the African and East Asian tales be classified as variants of ATU 333 or ATU 123?

205 Data for the study were drawn from 58 variants of ATU 333/123  
available in English translation from 33 populations (listed in Table  
S1). The tales comprise a representative sample of the geographic  
distribution of ATU 333/123 type tales (Figure 1), and the plot  
variations described in regional tale-type and motif indices [11]  
210 [41] [42] [44]. Relationships among the tales were reconstructed  
using three methods of phylogenetic analysis: cladistics, Bayesian  
inference and NeighbourNet (see Methods for a full description).  
The analyses focused on 72 plot variables, such as character  
of the protagonist (single child versus group of siblings; male  
215 versus female), the character of the villain (wolf, ogre, tiger,  
etc.), the tricks used by the villain to deceive the victim (false  
voice, disguised paws, etc.), whether the victim is devoured, escapes  
or is rescued, and so on. A full list of characters and explanation  
of the coding scheme is provided in the Supporting Information (File  
220 S1), together with the character matrix (File S2).