NBGP Cross-script Variant inclusion policy:

If, in any two given scripts, all the potential cross-script variants consist of dependent (e.g. Vowel Signs, Anusvara, Visarga, Chandrabindu etc.) characters **ONLY**, then that entire set can be ignored and no cross-script variants be proposed between those two scripts.

If, in any two given scripts, there is **AT LEAST ONE** non-dependent (e.g. Consonant, Vowel etc.) cross-script variant character/sequence present, all the potential cross-script variants be considered and proposed between the two scripts.

This cross-script analysis has been restricted to the scripts that have descended from the Brahmi as most of them share similar usage patterns. By and large, all of these scripts have a common set of characters that existed in Brahmi script and bear the same identities. However, as the scripts branched out from the Brahmi, depending on various factors, the shapes of the characters changed. This change in the shape was not uniform across all the characters and the scripts. Some characters shapes did change significantly whereas some of them still retained similarity. The cross-script similarity analysis also aims to identify such cases where the same character retained almost the same shape despite being part of the different scripts. These set of characters are variants of each other in true sense than merely of co-incidental visual similarity.

Case of Malayalam and Odia (Oriya) TTHA Consonant:
This is the case of "Consonant Ttha" which happened to retain the same shape despite being part of different scripts, i.e., Malayalam and Odia. These characters are:

ഠ - MALAYALAM LETTER TTHA (U+0D20)
ଠ - ORIYA LETTER TTHA (U+0B20)

Both the characters, look exactly alike and belong to a "Consonant" category. As they are consonants, each of them, even in the simplest form i.e. the characters themselves, are valid labels. As per the NBGP cross-script variant inclusion policy, this is a valid case for inclusion. Also, even if they are single characters, when the same character combines, theoretically they can form infinite[1] number of cross-script variant labels between the scripts involved. Here are some samples of some of those labels:

Malayalam	Oriya
ഠഠഠ U+0D20 U+0D20 U+0D20	ଠଠଠ U+0B20 U+0B20 U+0B20

---

[1] Though theoretically infinite, this number would be limited to the number of such labels whose equivalent punycode string would not exceed 63 characters including the ACE prefix "xn--".

OOOO U+0D20 U+0D20 U+0D20 U+0D20	OOOO U+0B20 U+0B20 U+0B20 U+0B20
OOOOO U+0D20 U+0D20 U+0D20 U+0D20 U+0D20	OOOOO U+0B20 U+0B20 U+0B20 U+0B20 U+0B20

Since, having such labels is a realistic possibility and the corresponding labels look almost exactly alike, NBGP has proposed them as blocked variants.

NBGP acknowledges the concern that this shape is quite generic and may have parallels in other scripts not under its ambit. However, as NBGP does not have any exposure about actual usage of those characters in those particular scripts, NBGP desisted from including them in the analysis. As NBGP has already considered all the related scripts under the cross-script variant analysis, the similarity of the characters belonging to NBGP scripts with other scripts not under the NBGP ambit, may be of a mere co-incidental visual nature.

Additionally, this concern is not limited to these two characters but for all the characters in all the scripts under the scope of the Root LGR procedure. Carrying out this analysis can practically be done only with the Generation Panels that exist while the NBGP is active. This still leaves out those scripts out of the scope which may not have a Generation Panel established yet. Hence, carrying out this exercise in entirety is quite impracticable. This conundrum can be resolved if all the such cases are handled by the "String Similarity Assessment Panel" of ICANN.