Generating Human Ortholog Lists from ENSEMBL BioMart

Load required libraries

```
library(plyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stringr)
library(biomaRt)
library(genefilter)
library(stringr)
library(devtools)
```

Purpose

To pull the current Ensembl annotations, including orthology type, for each of the genomes being used in this study.

```
##Looking at currently available databases in biomaRt
listMarts()
```

```
##                biomart               version
## 1 ENSEMBL_MART_ENSEMBL      Ensembl Genes 97
## 2   ENSEMBL_MART_MOUSE      Mouse strains 97
## 3     ENSEMBL_MART_SNP  Ensembl Variation 97
## 4 ENSEMBL_MART_FUNCGEN Ensembl Regulation 97
```

```
##Specifying that we want to work with the ENSEMBL database -- want to use ENSEMBL 96
##since this was the  version we used for processing our RNASeq reads.
ensembl <- useMart("ENSEMBL_MART_ENSEMBL",
          host = "http://apr2019.archive.ensembl.org",
          ensemblRedirect = FALSE)
```

```
## Warning in useMart("ENSEMBL_MART_ENSEMBL", host = "http://
## apr2019.archive.ensembl.org", : The argument "ensemblRedirect" has been
## deprecated and will be removed in the next biomaRt release.
```

```
datasets <- listDatasets(ensembl)
datasets
```

```
##                            dataset
## 1     abrachyrhynchus_gene_ensembl
## 2         acalliptera_gene_ensembl
## 3       acarolinensis_gene_ensembl
## 4        acitrinellus_gene_ensembl
## 5            ahaastii_gene_ensembl
## 6        amelanoleuca_gene_ensembl
## 7          amexicanus_gene_ensembl
## 8          anancymaae_gene_ensembl
## 9          aocellaris_gene_ensembl
## 10            aowenii_gene_ensembl
## 11           apercula_gene_ensembl
## 12     aplatyrhynchos_gene_ensembl
## 13      apolyacanthus_gene_ensembl
## 14              arowi_gene_ensembl
## 15       atestudineus_gene_ensembl
## 16             bbison_gene_ensembl
## 17             bmutus_gene_ensembl
## 18            btaurus_gene_ensembl
## 19        cabingdonii_gene_ensembl
## 20            caperea_gene_ensembl
## 21              catys_gene_ensembl
## 22            cbellii_gene_ensembl
## 23         ccaeruleus_gene_ensembl
## 24        ccanadensis_gene_ensembl
## 25         ccapucinus_gene_ensembl
## 26         cchok1gshd_gene_ensembl
## 27            ccrigri_gene_ensembl
## 28             cdingo_gene_ensembl
## 29          celegans_gene_ensembl
## 30        cfamiliaris_gene_ensembl
## 31            chircus_gene_ensembl
## 32         choffmanni_gene_ensembl
## 33      cintestinalis_gene_ensembl
## 34           cjacchus_gene_ensembl
## 35          cjaponica_gene_ensembl
## 36          clanigera_gene_ensembl
## 37         cpalliatus_gene_ensembl
## 38              cpicr_gene_ensembl
## 39         cporcellus_gene_ensembl
## 40           cporosus_gene_ensembl
## 41            cpugnax_gene_ensembl
## 42           cpygmaea_gene_ensembl
## 43           csabaeus_gene_ensembl
## 44          csavignyi_gene_ensembl
## 45        csemilaevis_gene_ensembl
## 46          csyrichta_gene_ensembl
## 47        cvariegatus_gene_ensembl
## 48      dmelanogaster_gene_ensembl
## 49   dnovaehollandiae_gene_ensembl
## 50      dnovemcinctus_gene_ensembl
```

```
## 51             dordii_gene_ensembl
## 52             drerio_gene_ensembl
## 53            easinus_gene_ensembl
## 54           eburgeri_gene_ensembl
## 55          ecaballus_gene_ensembl
## 56         eeuropaeus_gene_ensembl
## 57            elucius_gene_ensembl
## 58          etelfairi_gene_ensembl
## 59        falbicollis_gene_ensembl
## 60             fcatus_gene_ensembl
## 61        fdamarensis_gene_ensembl
## 62      fheteroclitus_gene_ensembl
## 63         gaculeatus_gene_ensembl
## 64           gaffinis_gene_ensembl
## 65         gagassizii_gene_ensembl
## 66            ggallus_gene_ensembl
## 67           ggorilla_gene_ensembl
## 68            gmorhua_gene_ensembl
## 69           hburtoni_gene_ensembl
## 70             hcomes_gene_ensembl
## 71            hfemale_gene_ensembl
## 72              hmale_gene_ensembl
## 73           hsapiens_gene_ensembl
## 74         ipunctatus_gene_ensembl
## 75  itridecemlineatus_gene_ensembl
## 76          jhyemalis_gene_ensembl
## 77           jjaculus_gene_ensembl
## 78        kmarmoratus_gene_ensembl
## 79          lafricana_gene_ensembl
## 80          lbergylta_gene_ensembl
## 81         lchalumnae_gene_ensembl
## 82          lcoronata_gene_ensembl
## 83         ldomestica_gene_ensembl
## 84          loculatus_gene_ensembl
## 85             malbus_gene_ensembl
## 86           marmatus_gene_ensembl
## 87           mauratus_gene_ensembl
## 88            mcaroli_gene_ensembl
## 89         mdomestica_gene_ensembl
## 90      mfascicularis_gene_ensembl
## 91              mfuro_gene_ensembl
## 92         mgallopavo_gene_ensembl
## 93       mleucophaeus_gene_ensembl
## 94         mlucifugus_gene_ensembl
## 95           mmarmota_gene_ensembl
## 96              mmola_gene_ensembl
## 97           mmulatta_gene_ensembl
## 98           mmurinus_gene_ensembl
## 99          mmusculus_gene_ensembl
## 100       mnemestrina_gene_ensembl
## 101      mochrogaster_gene_ensembl
## 102           mpahari_gene_ensembl
## 103       mspicilegus_gene_ensembl
## 104          mspretus_gene_ensembl
```

```
## 105        mundulatus_gene_ensembl
## 106     munguiculatus_gene_ensembl
## 107       mvitellinus_gene_ensembl
## 108            mzebra_gene_ensembl
## 109        nbrichardi_gene_ensembl
## 110          neugenii_gene_ensembl
## 111           ngalili_gene_ensembl
## 112       nleucogenys_gene_ensembl
## 113        nmeleagris_gene_ensembl
## 114       nperdicaria_gene_ensembl
## 115         nscutatus_gene_ensembl
## 116            nvison_gene_ensembl
## 117         oanatinus_gene_ensembl
## 118            oaries_gene_ensembl
## 119        ocuniculus_gene_ensembl
## 120            odegus_gene_ensembl
## 121        ogarnettii_gene_ensembl
## 122              ohni_gene_ensembl
## 123             ohsok_gene_ensembl
## 124          olatipes_gene_ensembl
## 125       omelastigma_gene_ensembl
## 126        oniloticus_gene_ensembl
## 127         oprinceps_gene_ensembl
## 128           pabelii_gene_ensembl
## 129          paltaica_gene_ensembl
## 130           panubis_gene_ensembl
## 131          pbairdii_gene_ensembl
## 132         pcapensis_gene_ensembl
## 133         pcinereus_gene_ensembl
## 134       pcoquereli_gene_ensembl
## 135          pformosa_gene_ensembl
## 136       pkingsleyae_gene_ensembl
## 137        platipinna_gene_ensembl
## 138   pmagnuspinnatus_gene_ensembl
## 139            pmajor_gene_ensembl
## 140          pmarinus_gene_ensembl
## 141         pmexicana_gene_ensembl
## 142       pnattereri_gene_ensembl
## 143         pnyererei_gene_ensembl
## 144         ppaniscus_gene_ensembl
## 145           ppardus_gene_ensembl
## 146       preticulata_gene_ensembl
## 147            psimus_gene_ensembl
## 148         psinensis_gene_ensembl
## 149     ptephrosceles_gene_ensembl
## 150      ptroglodytes_gene_ensembl
## 151         pvampyrus_gene_ensembl
## 152        pvitticeps_gene_ensembl
## 153            rbieti_gene_ensembl
## 154       rnorvegicus_gene_ensembl
## 155        rroxellana_gene_ensembl
## 156          saraneus_gene_ensembl
## 157      sboliviensis_gene_ensembl
## 158           scanaria_gene_ensembl
```

```
## 159       scerevisiae_gene_ensembl
## 160         sdauricus_gene_ensembl
## 161         sdorsalis_gene_ensembl
## 162         sdumerili_gene_ensembl
## 163         sformosus_gene_ensembl
## 164         sharrisii_gene_ensembl
## 165          smaximus_gene_ensembl
## 166         smerianae_gene_ensembl
## 167         spartitus_gene_ensembl
## 168        spunctatus_gene_ensembl
## 169           sscrofa_gene_ensembl
## 170        tbelangeri_gene_ensembl
## 171           tgelada_gene_ensembl
## 172          tguttata_gene_ensembl
## 173     tnigroviridis_gene_ensembl
## 174         trubripes_gene_ensembl
## 175        ttruncatus_gene_ensembl
## 176       uamericanus_gene_ensembl
## 177        umaritimus_gene_ensembl
## 178          uparryii_gene_ensembl
## 179            vpacos_gene_ensembl
## 180           vvulpes_gene_ensembl
## 181       xcouchianus_gene_ensembl
## 182        xmaculatus_gene_ensembl
## 183       xtropicalis_gene_ensembl
## 184       zalbicollis_gene_ensembl
##                                                        description
## 1                            Pink-footed goose genes (ASM259213v1)
## 2                                 Eastern happy genes (fAstCal1.2)
## 3                                   Anole lizard genes (AnoCar2.0)
## 4                                   Midas cichlid genes (Midas_v5)
## 5                                Great spotted kiwi genes (aptHaa1)
## 6                                            Panda genes (ailMel1)
## 7                     Mexican tetra genes (Astyanax_mexicanus-2.0)
## 8                                Ma's night monkey genes (Anan_2.0)
## 9                               Clown anemonefish genes (AmpOce1.0)
## 10                              Little spotted kiwi genes (aptOwe1)
## 11                                 Orange clownfish genes (Nemo_v1)
## 12                                         Duck genes (CAU_duck1.0)
## 13                               Spiny chromis genes (ASM210954v1)
## 14                              Okarito brown kiwi genes (aptRow1)
## 15                               Climbing perch genes (fAnaTes1.1)
## 16                              American bison genes (Bison_UMD1.0)
## 17                                   Wild yak genes (BosGru_v2.0)
## 18                                         Cow genes (ARS-UCD1.2)
## 19              Abingdon island giant tortoise genes (ASM359739v1)
## 20                           Brazilian guinea pig genes (CavAp1.0)
## 21                                Sooty mangabey genes (Caty_1.0)
## 22             Painted turtle genes (Chrysemys_picta_bellii-3.0.3)
## 23                                         Blue tit genes (cyaCae2)
## 24                         American beaver genes (C.can_genome_v1.0)
## 25                               Capuchin genes (Cebus_imitator-1.0)
## 26                     Chinese hamster CHOK1GS genes (CHOK1GS_HDv1)
## 27                        Chinese hamster CriGri genes (CriGri_1.0)
```

```
## 28                                    Dingo genes (ASM325472v1)
## 29                      Caenorhabditis elegans genes (WBcel235)
## 30                                       Dog genes (CanFam3.1)
## 31                                           Goat genes (ARS1)
## 32                                       Sloth genes (choHof1)
## 33                                  C.intestinalis genes (KH)
## 34                               Marmoset genes (ASM275486v1)
## 35                  Japanese quail genes (Coturnix_japonica_2.0)
## 36                      Long-tailed chinchilla genes (ChiLan1.0)
## 37                            Angola colobus genes (Cang.pa_1.0)
## 38                     Chinese hamster PICR genes (CriGri-PICR)
## 39                               Guinea Pig genes (Cavpor3.0)
## 40           Australian saltwater crocodile genes (CroPor_comp1)
## 41                                    Ruff genes (ASM143184v1)
## 42                    Spoon-billed sandpiper genes (ASM369795v1)
## 43                               Vervet-AGM genes (ChlSab1.1)
## 44                                C.savignyi genes (CSAV 2.0)
## 45                               Tongue sole genes (Cse_v1.0)
## 46                       Tarsier genes (Tarsius_syrichta-2.0.1)
## 47                    Sheepshead minnow genes (C_variegatus-1.0)
## 48                      Drosophila melanogaster genes (BDGP6.22)
## 49                                          Emu genes (droNov1)
## 50                                 Armadillo genes (Dasnov3.0)
## 51                               Kangaroo rat genes (Dord_2.0)
## 52                                    Zebrafish genes (GRCz11)
## 53                                  Donkey genes (ASM303372v1)
## 54                                 Hagfish genes (Eburgeri_3.2)
## 55                                       Horse genes (EquCab3.0)
## 56                                    Hedgehog genes (eriEur1)
## 57                                Northern pike genes (Eluc_V3)
## 58                         Lesser hedgehog tenrec genes (TENREC)
## 59                                 Flycatcher genes (FicAlb_1.4)
## 60                                   Cat genes (Felis_catus_9.0)
## 61                             Damara mole rat genes (DMR_v1.0)
## 62                 Mummichog genes (Fundulus_heteroclitus-3.0.2)
## 63                               Stickleback genes (BROAD S1)
## 64                      Western mosquitofish genes (ASM309773v1)
## 65                  Agassiz's desert tortoise genes (ASM289641v1)
## 66                                     Chicken genes (GRCg6a)
## 67                                   Gorilla genes (gorGor4)
## 68                                       Cod genes (gadMor1)
## 69                        Burton's mouthbrooder genes (AstBur1.0)
## 70                     Tiger tail seahorse genes (H_comes_QL1_v1)
## 71                Naked mole-rat female genes (HetGla_female_1.0)
## 72                         Naked mole-rat male genes (HetGla_1.0)
## 73                                     Human genes (GRCh38.p12)
## 74                           Channel catfish genes (IpCoco_1.2)
## 75                                   Squirrel genes (SpeTri2.0)
## 76                           Dark-eyed junco genes (ASM382977v1)
## 77                        Lesser Egyptian jerboa genes (JacJac1.0)
## 78                           Mangrove rivulus genes (ASM164957v1)
## 79                                   Elephant genes (Loxafr3.0)
## 80                               Ballan wrasse genes (BallGen_V1)
## 81                                   Coelacanth genes (LatCha1)
```

```
## 82         Blue-crowned manakin genes (Lepidothrix_coronata-1.0)
## 83                             Bengalese finch genes (LonStrDom1)
## 84                                   Spotted gar genes (LepOcu1)
## 85                                 Swamp eel genes (M_albus_1.0)
## 86                                Zig-zag eel genes (fMasArm1.1)
## 87                               Golden Hamster genes (MesAur1.0)
## 88                            Ryukyu mouse genes (CAROLI_EIJ_v1.1)
## 89                                        Opossum genes (monDom5)
## 90           Crab-eating macaque genes (Macaca_fascicularis_5.0)
## 91                                     Ferret genes (MusPutFur1.0)
## 92                                      Turkey genes (Turkey_2.01)
## 93                                       Drill genes (Mleu.le_1.0)
## 94                                     Microbat genes (Myoluc2.0)
## 95                                  Alpine marmot genes (marMar2.1)
## 96                                Ocean sunfish genes (ASM169857v1)
## 97                                       Macaque genes (Mmul_8.0.1)
## 98                                    Mouse Lemur genes (Mmur_3.0)
## 99                                         Mouse genes (GRCm38.p6)
## 100                             Pig-tailed macaque genes (Mnem_1.0)
## 101                                  Prairie vole genes (MicOch1.0)
## 102                             Shrew mouse genes (PAHARI_EIJ_v1.1)
## 103                                    Steppe mouse genes (MUSP714)
## 104                             Algerian mouse genes (SPRET_EiJ_v1)
## 105                 Budgerigar genes (Melopsittacus_undulatus_6.3)
## 106                         Mongolian gerbil genes (MunDraft-v1.0)
## 107                    Golden-collared manakin genes (ASM171598v2)
## 108                               Zebra mbuna genes (M_zebra_UMD2a)
## 109                              Lyretail cichlid genes (NeoBri1.0)
## 110                                       Wallaby genes (Meug_1.0)
## 111 Upper Galilee mountains blind mole rat genes (S.galili_v1.0)
## 112                                        Gibbon genes (Nleu_3.0)
## 113                           Helmeted guineafowl genes (NumMel1.0)
## 114                                 Chilean tinamou genes (notPer1)
## 115                        Mainland tiger snake genes (TS10Xv2-PRI)
## 116                                 American mink genes (NNQGG.v01)
## 117                                          Platypus genes (OANA5)
## 118                                          Sheep genes (Oar_v3.1)
## 119                                       Rabbit genes (OryCun2.0)
## 120                                         Degu genes (OctDeg1.0)
## 121                                       Bushbaby genes (OtoGar3)
## 122                          Japanese medaka HNI genes (ASM223471v1)
## 123                         Japanese medaka HSOK genes (ASM223469v1)
## 124                         Japanese medaka HdrR genes (ASM223467v1)
## 125                               Indian medaka genes (Om_v0.7.RACA)
## 126                                  Tilapia genes (Orenil1.0)
## 127                                 Pika genes (OchPri2.0-Ens)
## 128                                    Orangutan genes (PPYG2)
## 129                                      Tiger genes (PanTig1.0)
## 130                                Olive baboon genes (Panu_3.0)
## 131              Northern American deer mouse genes (HU_Pman_2.1)
## 132                                        Hyrax genes (proCap1)
## 133                                Koala genes (phaCin_unsw_v4.1)
## 134                             Coquerel's sifaka genes (Pcoq_1.0)
## 135                    Amazon molly genes (Poecilia_formosa-5.1.2)
```

```
## 136                  Paramormyrops kingsleyae genes (PKINGS_0.1)
## 137                        Sailfin molly genes (P_latipinna-1.0)
## 138                  Periophthalmus magnuspinnatus genes (PM.fa)
## 139                              Great Tit genes (Parus_major1.1)
## 140                                  Lamprey genes (Pmarinus_7.0)
## 141                           Shortfin molly genes (P_mexicana-1.0)
## 142      Red-bellied piranha genes (Pygocentrus_nattereri-1.0.2)
## 143                          Makobe Island cichlid genes (PunNye1.0)
## 144                                       Bonobo genes (panpan1.1)
## 145                                     Leopard genes (PanPar1.0)
## 146                               Guppy genes (Guppy_female_1.0_MT)
## 147                          Greater bamboo lemur genes (Prosim_1.0)
## 148                  Chinese softshell turtle genes (PelSin_1.0)
## 149                          Ugandan red Colobus genes (ASM277652v2)
## 150                                Chimpanzee genes (Pan_tro_3.0)
## 151                                      Megabat genes (pteVam1)
## 152                            Central bearded dragon genes (pvi1.1)
## 153                  Black snub-nosed monkey genes (ASM169854v1)
## 154                                       Rat genes (Rnor_6.0)
## 155                         Golden snub-nosed monkey genes (Rrox_v1)
## 156                                      Shrew genes (sorAra1)
## 157                       Bolivian squirrel monkey genes (SaiBol1.0)
## 158                                    Common canary genes (SCA1)
## 159                         Saccharomyces cerevisiae genes (R64-1-1)
## 160                  Daurian ground squirrel genes (ASM240643v1)
## 161                          Yellowtail amberjack genes (Sedor1)
## 162                            Greater amberjack genes (Sdu_1.0)
## 163                          Asian bonytongue genes (ASM162426v1)
## 164                         Tasmanian devil genes (Devil_ref v7.0)
## 165                                    Turbot genes (ASM318616v1)
## 166               Argentine black and white tegu genes (HLtupMer3)
## 167            Bicolor damselfish genes (Stegastes_partitus-1.0.2)
## 168                                  Tuatara genes (ASM311381v1)
## 169                                      Pig genes (Sscrofa11.1)
## 170                                   Tree Shrew genes (tupBel1)
## 171                                      Gelada genes (Tgel_1.0)
## 172                                Zebra Finch genes (taeGut3.2.4)
## 173                                Tetraodon genes (TETRAODON 8.0)
## 174                                           Fugu genes (FUGU5)
## 175                                      Dolphin genes (turTru1)
## 176                          American black bear genes (ASM334442v1)
## 177                                    Polar bear genes (UrsMar_1.0)
## 178                       Arctic ground squirrel genes (ASM342692v1)
## 179                                       Alpaca genes (vicPac1)
## 180                                     Red fox genes (VulVul2.2)
## 181       Monterrey platyfish genes (Xiphophorus_couchianus-4.0.1)
## 182                           Platyfish genes (X_maculatus-5.0-male)
## 183                                      Xenopus genes (JGI 4.2)
## 184  White-throated sparrow genes (Zonotrichia_albicollis-1.0.1)
##                          version
## 1                      ASM259213v1
## 2                       fAstCal1.2
## 3                        AnoCar2.0
## 4                         Midas_v5
```

```
## 5                        aptHaa1
## 6                        ailMel1
## 7         Astyanax_mexicanus-2.0
## 8                         Anan_2.0
## 9                        AmpOce1.0
## 10                       aptOwe1
## 11                       Nemo_v1
## 12                    CAU_duck1.0
## 13                    ASM210954v1
## 14                       aptRow1
## 15                     fAnaTes1.1
## 16                   Bison_UMD1.0
## 17                    BosGru_v2.0
## 18                     ARS-UCD1.2
## 19                    ASM359739v1
## 20                       CavAp1.0
## 21                       Caty_1.0
## 22  Chrysemys_picta_bellii-3.0.3
## 23                       cyaCae2
## 24              C.can_genome_v1.0
## 25             Cebus_imitator-1.0
## 26                   CHOK1GS_HDv1
## 27                     CriGri_1.0
## 28                    ASM325472v1
## 29                       WBcel235
## 30                     CanFam3.1
## 31                          ARS1
## 32                       choHof1
## 33                            KH
## 34                   ASM275486v1
## 35          Coturnix_japonica_2.0
## 36                     ChiLan1.0
## 37                    Cang.pa_1.0
## 38                    CriGri-PICR
## 39                     Cavpor3.0
## 40                   CroPor_comp1
## 41                    ASM143184v1
## 42                    ASM369795v1
## 43                     ChlSab1.1
## 44                      CSAV 2.0
## 45                       Cse_v1.0
## 46        Tarsius_syrichta-2.0.1
## 47              C_variegatus-1.0
## 48                        BDGP6.22
## 49                       droNov1
## 50                     Dasnov3.0
## 51                      Dord_2.0
## 52                        GRCz11
## 53                    ASM303372v1
## 54                   Eburgeri_3.2
## 55                      EquCab3.0
## 56                       eriEur1
## 57                       Eluc_V3
## 58                        TENREC
```

```
## 59                    FicAlb_1.4
## 60                Felis_catus_9.0
## 61                       DMR_v1.0
## 62   Fundulus_heteroclitus-3.0.2
## 63                       BROAD S1
## 64                    ASM309773v1
## 65                    ASM289641v1
## 66                         GRCg6a
## 67                       gorGor4
## 68                       gadMor1
## 69                     AstBur1.0
## 70                 H_comes_QL1_v1
## 71              HetGla_female_1.0
## 72                     HetGla_1.0
## 73                     GRCh38.p12
## 74                     IpCoco_1.2
## 75                      SpeTri2.0
## 76                    ASM382977v1
## 77                      JacJac1.0
## 78                    ASM164957v1
## 79                      Loxafr3.0
## 80                     BallGen_V1
## 81                        LatCha1
## 82      Lepidothrix_coronata-1.0
## 83                     LonStrDom1
## 84                        LepOcu1
## 85                    M_albus_1.0
## 86                     fMasArm1.1
## 87                     MesAur1.0
## 88                CAROLI_EIJ_v1.1
## 89                        monDom5
## 90       Macaca_fascicularis_5.0
## 91                   MusPutFur1.0
## 92                    Turkey_2.01
## 93                    Mleu.le_1.0
## 94                      Myoluc2.0
## 95                      marMar2.1
## 96                    ASM169857v1
## 97                     Mmul_8.0.1
## 98                       Mmur_3.0
## 99                      GRCm38.p6
## 100                      Mnem_1.0
## 101                     MicOch1.0
## 102              PAHARI_EIJ_v1.1
## 103                       MUSP714
## 104                  SPRET_EiJ_v1
## 105  Melopsittacus_undulatus_6.3
## 106                MunDraft-v1.0
## 107                  ASM171598v2
## 108                M_zebra_UMD2a
## 109                      NeoBri1.0
## 110                       Meug_1.0
## 111                S.galili_v1.0
## 112                       Nleu_3.0
```

```
## 113                   NumMel1.0
## 114                     notPer1
## 115                  TS10Xv2-PRI
## 116                    NNQGG.v01
## 117                        OANA5
## 118                     Oar_v3.1
## 119                    OryCun2.0
## 120                    OctDeg1.0
## 121                     OtoGar3
## 122                  ASM223471v1
## 123                  ASM223469v1
## 124                  ASM223467v1
## 125                 Om_v0.7.RACA
## 126                    Orenil1.0
## 127                 OchPri2.0-Ens
## 128                         PPYG2
## 129                    PanTig1.0
## 130                     Panu_3.0
## 131                  HU_Pman_2.1
## 132                     proCap1
## 133             phaCin_unsw_v4.1
## 134                     Pcoq_1.0
## 135       Poecilia_formosa-5.1.2
## 136                   PKINGS_0.1
## 137              P_latipinna-1.0
## 138                        PM.fa
## 139               Parus_major1.1
## 140                 Pmarinus_7.0
## 141               P_mexicana-1.0
## 142  Pygocentrus_nattereri-1.0.2
## 143                    PunNye1.0
## 144                    panpan1.1
## 145                    PanPar1.0
## 146          Guppy_female_1.0_MT
## 147                   Prosim_1.0
## 148                   PelSin_1.0
## 149                  ASM277652v2
## 150                  Pan_tro_3.0
## 151                      pteVam1
## 152                       pvi1.1
## 153                  ASM169854v1
## 154                     Rnor_6.0
## 155                     Rrox_v1
## 156                     sorAra1
## 157                    SaiBol1.0
## 158                         SCA1
## 159                     R64-1-1
## 160                  ASM240643v1
## 161                       Sedor1
## 162                     Sdu_1.0
## 163                  ASM162426v1
## 164               Devil_ref v7.0
## 165                  ASM318616v1
## 166                    HLtupMer3
```

```
## 167     Stegastes_partitus-1.0.2
## 168                   ASM311381v1
## 169                   Sscrofa11.1
## 170                       tupBel1
## 171                     Tgel_1.0
## 172                   taeGut3.2.4
## 173                TETRAODON 8.0
## 174                         FUGU5
## 175                       turTru1
## 176                   ASM334442v1
## 177                    UrsMar_1.0
## 178                   ASM342692v1
## 179                       vicPac1
## 180                     VulVul2.2
## 181 Xiphophorus_couchianus-4.0.1
## 182         X_maculatus-5.0-male
## 183                       JGI 4.2
## 184 Zonotrichia_albicollis-1.0.1
```

```
##Pull in Ensembl marts for all species we used in this study
pa_ensembl <- useDataset("pabelii_gene_ensembl", mart = ensembl)
saimiri_ensembl <- useDataset("sboliviensis_gene_ensembl", mart = ensembl)
chimp_ensembl <- useDataset("ptroglodytes_gene_ensembl", mart = ensembl)
bonobo_ensembl <- useDataset("ppaniscus_gene_ensembl", mart = ensembl)
mn_ensembl <- useDataset("mnemestrina_gene_ensembl", mart = ensembl)
bab_ensembl <- useDataset("panubis_gene_ensembl", mart = ensembl)
gorilla_ensembl <- useDataset("ggorilla_gene_ensembl", mart = ensembl)
rhmac_ensembl <- useDataset("mmulatta_gene_ensembl", mart = ensembl)
human_ensembl <- useDataset("hsapiens_gene_ensembl", mart = ensembl)

##Making list of the above Ensembl marts
ensembl_marts <- list(bab_ensembl, bonobo_ensembl, chimp_ensembl,
                      gorilla_ensembl, human_ensembl, pa_ensembl, mn_ensembl,  rhmac_ensembl,
                      saimiri_ensembl)
names(ensembl_marts) <- c("bab", "bonobo", "chimp", "gorilla", "human", "orang", "ptmac",
                          "rhesus", "sqmonk")


##Building our biomaRt query (used function listAttributes for a given mart
##initially in order to see what we can pull in from biomaRt)
##"featurepage" function will allow us to get the description, gene name, and the biotype for
##every ENSEMBL ID listed for a given species.
featurepage <- function(species_ensembl) {
  getBM(attributes = c('ensembl_gene_id', 'description',
                               'external_gene_name', 'gene_biotype'),
                  mart = species_ensembl)
}

##Homologs come from another "page" of attributes so this function separately pulls in that
##information which we could not get from our "featurepage" function above. This lists for
##all the ENSEMBL IDs for a given species the ENSEMBL ID and gene name for human homolog(s)
##(if there is one), and the type of ortholog (one-to-one, one-to-many, etc.) it is.
homologypage <- function(species_ensembl) {
  getBM(attributes = c('ensembl_gene_id',
```

```
                          'hsapiens_homolog_ensembl_gene',
                          'hsapiens_homolog_orthology_type',
                          'hsapiens_homolog_associated_gene_name'),
                         mart = species_ensembl)
}

##Now applying the "featurepage" function to the list of ENSEMBL IDs (the first column of each
##data frame contained in the "NHP_geno_df" list) for each NHP species.
##Performing the function on only the non-human primate samples.
feature_species <- llply(ensembl_marts, featurepage)
save(feature_species, file = paste0(Sys.Date(), "genomefeatures_species.Rdata"))

##Now applying the "homologypage" function to identify human homologs for each of the NHP genes.
##We do not need to include human
names(ensembl_marts)
```

```
## [1] "bab"     "bonobo"  "chimp"   "gorilla" "human"   "orang"   "ptmac"
## [8] "rhesus"  "sqmonk"
```

```
homology_species <- llply(ensembl_marts[c(1:4, 6:9)], homologypage)

save(homology_species, file = paste0(Sys.Date(), "homology_species.Rdata"))

##Combining the information from the "feature_species"" and "homology_species" outputs to have all
##the information together in one data frame for each species. Since we excluded human from the
##homology_species, since we can't get homologs for human homologs, we need to limit the
##feature_species to the same ones we have in the homology_species.
names(homology_species)
```

```
## [1] "bab"     "bonobo"  "chimp"   "gorilla" "orang"   "ptmac"   "rhesus"
## [8] "sqmonk"
```

```
names(feature_species)
```

```
## [1] "bab"     "bonobo"  "chimp"   "gorilla" "human"   "orang"   "ptmac"
## [8] "rhesus"  "sqmonk"
```

```
homology_feature <- mapply(merge, homology_species, feature_species[c(1:4, 6:9)],
            SIMPLIFY = FALSE, USE.NAMES = TRUE)
save(homology_feature, file = paste0(Sys.Date(), "homology_feature_NHP.Rdata"))
str(homology_feature[1])
```

```
## List of 1
##  $ bab:'data.frame': 31891 obs. of  7 variables:
##   ..$ ensembl_gene_id                     : chr [1:31891] "ENSPANG00000000039" "ENSPANG00000000040"
##   ..$ hsapiens_homolog_ensembl_gene       : chr [1:31891] "ENSG00000275132" "ENSG00000136560" "ENSG
##   ..$ hsapiens_homolog_orthology_type     : chr [1:31891] "ortholog_one2one" "ortholog_one2one" "or
##   ..$ hsapiens_homolog_associated_gene_name: chr [1:31891] "RN7SL663P" "TANK" "PSMD14" "" ...
##   ..$ description                         : chr [1:31891] "RNA, 7SL, cytoplasmic 663, pseudogene [S
##   ..$ external_gene_name                  : chr [1:31891] "RF00017" "TANK" "PSMD14" "RF00026" ...
##   ..$ gene_biotype                        : chr [1:31891] "misc_RNA" "protein_coding" "protein_codi
```

```
one2ones <- llply(homology_feature, function(x) dplyr::filter(x,
                  hsapiens_homolog_orthology_type == "ortholog_one2one"))
save(one2ones, file = paste0(Sys.Date(), "one2oneorthologs_allspecies.Rdata"))
```

```
one2ones_common <- Reduce(intersect, llply(one2ones, function(x) x$hsapiens_homolog_ensembl_gene))
save(one2ones_common, file = paste0(Sys.Date(), "one2oneorthologs_common_allspecies.Rdata"))
```

Session Info

```
sessionInfo()
```

```
## R version 3.5.2 (2018-12-20)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] bindrcpp_0.2.2    usethis_1.4.0     devtools_2.0.1    genefilter_1.64.0
## [5] biomaRt_2.38.0    stringr_1.3.1     dplyr_0.7.8       plyr_1.8.4
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.0           lattice_0.20-38      prettyunits_1.0.2
##  [4] ps_1.3.0             assertthat_0.2.0     rprojroot_1.3-2
##  [7] digest_0.6.18        R6_2.3.0             backports_1.1.3
## [10] stats4_3.5.2         RSQLite_2.1.1        evaluate_0.12
## [13] httr_1.4.0           pillar_1.3.1         rlang_0.3.1
## [16] progress_1.2.0       curl_3.3             annotate_1.60.0
## [19] callr_3.1.1          blob_1.1.1           S4Vectors_0.20.1
## [22] Matrix_1.2-15        rmarkdown_1.11       desc_1.2.0
## [25] splines_3.5.2        RCurl_1.95-4.11      bit_1.1-14
## [28] compiler_3.5.2       xfun_0.4             pkgconfig_2.0.2
## [31] BiocGenerics_0.28.0  pkgbuild_1.0.2       htmltools_0.3.6
## [34] tidyselect_0.2.5     tibble_2.0.1         IRanges_2.16.0
## [37] XML_3.98-1.16        crayon_1.3.4         withr_2.1.2
## [40] bitops_1.0-6         grid_3.5.2           xtable_1.8-3
## [43] DBI_1.0.0            magrittr_1.5         cli_1.0.1
## [46] stringi_1.2.4        fs_1.2.6             remotes_2.0.2
## [49] tools_3.5.2          bit64_0.9-7          Biobase_2.42.0
## [52] glue_1.3.0           purrr_0.2.5          hms_0.4.2
## [55] processx_3.2.1       pkgload_1.0.2        parallel_3.5.2
## [58] survival_2.43-3      yaml_2.2.0           AnnotationDbi_1.44.0
## [61] sessioninfo_1.1.1    memoise_1.1.0        knitr_1.21
## [64] bindr_0.1.1
```