PHYLOGENOMIC ANALYSIS OF THE METABOTROPIC P2Y RECEPTOR FAMILY AND ITS EXPRESSION IN ZEBRAFISH, *Danio rerio*

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Yen Yen Kwan

aus Kuala Lumpur, Malaysia

Köln 2008

Berichterstatter: Prof. Dr. S. I. Korsching

Prof. Dr. Jens Brüning

Tag der mundlichen Prufung: 25 November 2008

ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen in der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfungvorgelegen hat; dass sie - abgesehen von den unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist, sowie dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. S. I. Korsching betreut worden.

Köln, den 8. Oktober, 2008

1. ACKNOWLEDGEMENTS

Firstly, I would like to thank Prof Sigrun Korshing for her guidance, patience, advice and support throughout these four years. Besides that, I would also like to express my appreciation to Prof Dr Jens Brüning, and Prof Dr Peter Kloppenburg for being in my thesis committee.

Next, I would like to acknowledge the International Graduate School in Genetics and Functional Genomics for providing the fellowship of my study. Special thanks to Dr Brigitte Wilcken-Bergmann for her help and support in all these years. Thank you very much.

Coming to Germany was a wonderful chapter in my life. I had my own fair share of good times here. To Joanna, Daniela, Luis and Leo, thank you for your comradeship, motivation, and support for the good and bad moments in all these years. Many thanks to Hayati, Hua, Ying, Jinu, Channa and many others who are not listed here for their friendship and hospitability

Many thanks to my labmates for their support, help and encouragement. Special thanks extended to Dr Yuichiro Oka and Luis Saraiva for their fruitful discussions, criticisms and suggestions. I would like to thank Heven Midani for her pleasant coffee-time company and Mehmet Saltürk for taking care of the lab.

To my buddies from Malaysia, Suky, Mitty, Mei Fung, Ai Bee and Yen Ling, my sincere thanks for your constant support for all the ups and downs, joy and sadness, laughters and tears. Many years have passed and our friendship is still blooming, and I hope our friendship will remain the same in the coming years.

To my parents, no words can describe my gratitude towards you. Both of you have been the pillar of strength in my life. In you, I learned about empathy, conscience, filiality and forgiveness. Thank you for your unlimited love, care and support even at all times To my siblings, it's fortunate to be with you. My dearest sisters for their love and encouragement. My dear brothers for their kindness and source of motivation. Last but not least, my aunt Thong Loy, for her motherly comforts.

II. TABLE OF CONTENTS

III. LIST OF FIGURES

IV. LIST OF TABLES

V. ABSTRACT

G-protein-coupled receptors (GPCRs) constitute the largest and most divergent class of cell surface proteins. GPCRs can be activated by extracellular signals as diverse as light, peptides, proteins, lipids, odorants, tastants, nucleotides and nucleosides. Currently, all the known chemosensory receptor genes, such as odorant, taste and pheromone receptors belong to the GPCR family, with many of them being class A or class-A related genes. Metabotropic P2Y receptors belong to class A GPCRs and are activated by extracellular nucleotides. Nucleotides are released to the environment when the organism is injured and therefore serve as one of the food stimuli. As such, it would be interesting to see if P2Y receptors play any role in olfaction in zebrafish.

To date, eight functional human P2Y receptors and 25 related orphan receptors have been found. I performed extensive data mining in drosophila, ascidian, jawless, cartilaginous and bony fish, frog and human genomes to delineate the P2Y family and investigate its evolutionary origin. The P2Y family originates early in the vertebrate lineage, reflected by the presence of lamprey and the absence of ascidian orthologues. Consistent with these findings, no P2Y receptor is found in invertebrates. In total, 38 subfamilies can be distinguished within the P2Y family, at least two third of which are already present in the shark genome. Two subfamilies, p2yl-3 and p2yl-4, are lost in the human lineage and only GPR 87 subfamily is lost in all teleost species.

Zebrafish has 68 P2Y receptor genes, the most of any fish species, and almost double as many than mammals. The teleost P2Y genes are widely distributed in the genome as small cluster and singletons. The vast majority of P2Y genes are intronless while the remaining genes contain up to five introns. In the teleost lineage, the genomic arrangement of P2Y genes is preserved to a large extent, and some synteny is found even with the elephant shark and human genome, possibly reflecting the functional importance of these genes. Selective pressure on teleost P2Y genes generally is high, as evidenced by a preponderance of negative selection. However, a few genes exhibit positive selection at individual sites.

In early development, P2Y genes are expressed in many tissues and organs, notably the central and peripheral nervous system, pharyngeal arches, otic vesicle and kidney, suggesting an important role in the development of many tissues. However, no expression is detected in larval olfactory epithelium. In contrast, olfactory epithelium of adult fish does express several P2Y genes as shown by RT-PCR. A possible explanation would be a late onset of expression in the olfactory epithelium. In situ hybridisation of adult olfactory epithelium established an ubiquitous distribution, both in the sensory and non-sensory region, which seems to argue against a role of P2Y genes in nucleotide odor detection. However, further studies will be necessary to give a definitive answer to that question.

VI. ZUSAMMENFASSUNG

Bei den G-protein-gekoppelten Rezeptoren (GPCRs) handelt es sich um die größte und vielfältigste Gruppe von Membranproteinen. GPCRs werden durch sehr diverse extrazelluläre Signale aktiviert, von Licht über Peptide, Proteine zu Geruchs- und Geschmacksstoffen, bis hin zu Lipiden, Nukleotiden und Nukleosiden. Alle gegenwärtig bekannten chemosensorischen Rezeptoren, wie die Geruchs-, Geschmacks- und Pheromonrezeptoren gehören der GPCR Familie an, und viele davon sind aus der Klasse A der GPCR. Das gilt auch für die metabotropen P2Y Rezeptoren, die durch extrazelluläre Nukleotide aktiviert werden. Nukleotide werden freigesetzt, wenn ein Organismus verletzt wird, und können so als Signale zur Nahrungsdetektion dienen. Es erschien daher interessant, zu untersuchen, ob die von Fischen bekannte Geruchsdetektion von Nukleotiden durch P2Y Rezeptoren vermittelt sein könnte.

Die P2Y Familie im Menschen besteht aus acht funktionalen P2Y Rezeptoren und 25 verwandten Rezeptoren mit bisher unbekannten Liganden. Ich habe ausführliche Datenbanksuchen in Genomen der Taufliege, von Aszidien, Kieferlosen, Knorpelfische und Knochenfische, sowie in Frosch und dem menschlichen Genom durchgeführt, um die P2Y Familie zu identifizieren und von benachbarten Familien abzugrenzen. Meine Ergebnisse zeigen, dass die P2Y Familie in frühen Wirbeltieren entstanden ist, da Vertreter dieser Familie im Neunauge, aber nicht in Aszidien vorkommen. Konsistent zu diesen Befunden konnten auch keine P2Y Rezeptoren in Wirbellosen gefunden werden. Insgesamt können 38 Unterfamilien innerhalb der P2Y Familie unterschieden werden, von denen mindestens 2/3 bereits im Haifischgenom vorhanden sind. Zwei Unterfamilien, p2yl-3 und p2yl-4, sind in der menschlichen Entwicklungslinie verloren gegangen, eine weitere Unterfamilie, GPR87, ist am Beginn der Knochenfischentwicklung verloren gegangen.

Der Zebrabärbling weist 68 P2Y Gene auf, am meisten von allen Fischarten, und fast doppelt so viele wie die Säuger. Die P2Y Gene kommen breit verteilt als kleine Cluster oder Einzelgene im Fischgenom vor. Die überwiegende Zahl der P2Y Rezeptorgene hat keine Introns, manche jedoch weisen bis zu fünf Introns auf. In der Entwicklungslinie der Knochenfische ist die relative genomische Anordnung der P2Y Gene zum großen Teil erhalten, und eine gewisse Syntenie läßt sich sogar zu den Genomen von Haifisch und Mensch beobachten. Eine solch hohe Konservierung spricht für die funktionale Bedeutung dieser Gene. Der Selektionsdruck auf die P2Y Gene der Knochenfische ist hoch, wie aus dem Überwiegen negativer Selektion ersichtlich wird. Manche Gene zeigen allerdings positive Selektion an vereinzelten Positionen.

In der frühen Entwicklung werden P2Y Gene in vielen Geweben und Organen exprimiert, insbesondere dem peripheren und zentralen Nervensystem, den Kiemenbögen, dem Ohrvesikel und der Niere, konsistent mit einer vielfältigen Rolle dieser Gene in der Entwicklung vieler Gewebe. Allerdings liess sich keine Expression im larvalen Riechepithel feststellen. Im Gegensatz dazu werden viele P2Y Gene im adulten Riechepithel exprimiert, wie ich sowohl mit RT-PCR als auch In situ-Hybridisierung zeigen konnte. Eine mögliche Erklärung dieser Diskrepanz wäre ein später Beginn der Expression im Riechepithel. In der In situ-Hybridisierung zeigt sich eine breite Expression der P2Y Gene sowohl im sensorischen als auch dem nichtsensorischen Bereich des Riechepithels, unerwartet für eine mögliche Funktion der P2Y Gene in der Geruchsdetektion von Nukleotiden. Weitere Studien werden erforderlich sein, um diese Frage abschliessend zu klären.

1. Introduction

A family of nucleotide receptor in teleost fishes, P2Y, is described in this thesis. As detailed below, there is a reason to suggest a possible function as sensory receptors for members of this family. This work presented here serve as a basis to investigate this possibility.

1.1 Nucleotides

Nucleotides are molecules composed of a nitrogen-containing ring compound linked to a five-carbon sugar and a phosphate group. The sugar involved in the synthesis and structure of a nucleotide can be either ribose or deoxyribose, Ribonucleotides are nucleotides containing ribose whereas deoxynucleotides are nucleotides containing deoxyribose. The phosphate group may consist of up to three phosphates forming monophosphates, diphosphates, or triphosphates, respectively. Nucleotide is named after the utilized base, either purine or pyrimidine. Purine bases are adenine and guanine whereas pyrimidines are cytosine, thymine and uracil.

Nucleotides are essential in the cell. They are the building blocks of the hereditary information. Besides that, nucleotides have many other functions in the cell: i) carrier of chemical energy in the cell, ii) coenzyme when combine with vitamin or its derivative and iii) signalling molecule in the cell. Ribonucleotide adenosine triphosphate, ATP, is used to transfer energy in hundreds of different cellular reactions. ATP is synthesized via the addition of a phosphate group to ADP. When energy is required, ATP releases its energy through its hydrolysis to ADP and inorganic phosphate. The regenerated ADP is reused for subsequent phosphorylation reactions that form ATP. Nucleotides bind to vitamin and its derivatives to form coenzyme. Coenzyme involved in both group-transfer and redox reactions in metabolism. Cyclic AMP (cAMP) is one of the signaling molecules in the cell. cAMP is derived from ATP by plasma membrane bound enzyme adenylnyl cyclase and is involved in intracellular signal transduction in organism.

1.2 Purine and pyrimidine

Purine consist of a pyrimidine ring (six-membered ring) fused to imidazole ring (five-membered ring). Members of purine include adenine, guanine, hypoxanthine, xanthine, theobromine, caffeine, uric acid and isoguanine. Purines are involved in many aspects of cellular processes: intermediary metabolism, nucleic acid synthesis and the supply of energy phosphates to various active transport systems. In cells, ATP is abundantly found as ATP is the main energy-providing

and energy-storing molecule for all biological processes. On the other hand, GTP is rarely used as an energy donor but rather to transmit signals throughout the cell notably in G-protein signaling cascade. In additions, purines appear to be ligand of specific receptor reside within plasma membrane which mediate changes in cell function in response to extracellular purines (Stone, 1982).

Unlike purine, pyrimidine comprise of six-membered ring. Pyrimidine members include cytosine, thymine and uracil. In the cell, pyrimidine serves as energy source for specific metabolic reactions. UTP is the source for activating glucose and galactose whereas CTP is an energy source in lipid metabolism.

1.3 Purine receptor

Purine receptors are cell surface receptors that transduce extracellular signals to an intracellular signal to induce a cellular response. There are two main families of purine receptors, P1 and P2 receptors. The P is an abbreviation for purinergic. P1 receptor is activated by adenosine whereas P2 receptor is activated by nucleotides (Abbracchio *et al*., 2006).

P1 or adenosine receptor family comprises of A_1, A_{2A}, A_{2B}, and A_3 adenosine receptors. The adenosine receptor family belongs to G- protein coupled receptor (GPCR). All the receptors couple to G proteins to activate the signaling cascades. A_1 receptor couples to both $G_{i/o}$ and $G_{q/11}$ to activate phospholipase C pathway whereas A_1 receptor only couples to $G_{i/o}$. On the other hand, both A_{2A} and A_{2B} receptors couple to G_s to increase cAMP concentration in the cell (Ralevic and Burnstock, 1998). Adenosine receptors are involved in modulation of cardiovascular, immune and central nervous systems (Burnstock, 2007).

P2 receptors are divided into two major families, P2X and P2Y. P2X family is of ligand-gated ion channel whereas P2Y family belongs to G-protein-coupled receptor (GPCR). P2X family is activated by extracellular adenosine 5'-triphosphate (ATP) and seven subunits have been identified to date, $P2X_{1-7}$. All the members possess two transmembrane-spanning regions in which N and C termini that have consensus-binding motifs for protein kinases reside intracellularly. The first TM is involved in channel gating and the second lining the ion pore (Fields and Burnstock, 2006). P2X receptors are homomeric or heteromeric assemblies of three subunits (Koles *et al*., 2008) P2X receptors mediate rapid (onset within 10ms) non-selective passage of cations (Na^+, K^+, Ca^{2+}) across the cell membrane resulting in an increase of Ca^{2+}

and depolarization (Bean, 1992; Dubyak and el-Moatassim, 1993). The direct influx of extracellular Ca^{2+} through the channel constitutes a significant source of the increase in intracellular Ca^{2+}. However, membrane depolarization leads to the secondary activation of voltage-dependent Ca^{2+} channels, which probably make the primary contribution to Ca^{2+} influx and to the increase in intracellular Ca^{2+}. Since this transduction mechanism does not depend on the production and diffusion of second-messengers within the cytosol or cell membrane, the response time is very rapid, and appropriately plays an important role in fast neuronal signaling and regulation of muscle contractility (Ralevic and Burnstock, 1998). P2X receptors are important in neuronal signaling, pain transmission and inflammation.

1.4 P2Y receptor

P2Y receptors are cell-surface receptors that belong to G-protein coupled receptor (GPCR) superfamily. As such, P2Y receptors share the common molecular architecture of GPCR, consisting of seven transmembrane domains connected by three intracellular and three extracellular loops. GPCR superfamily composed of five families, namely rhodopsin, secretin, adhesion, glutamate and frizzled/taste2. P2Y receptors belong to rhodopsin/Class A family, which is the largest family of GPCRs. Rhodopsin family of GPCRs have short N termini and the receptors are primarily activated by interactions of ligand with transmembrane regions and extracellular loops. This metabotropic receptor form the δ-group together with MAS-related, glycoprotein and olfactory receptors in rhodopsin family. P2Y receptors are activated not only by ATP, but also naturally occurring nucleotides or nucleotide sugars such as ADP, UTP, UDP and UDP glucose (Fredriksson *et al*., 2003; Abbracchio *et al*., 2006; Langerstrom and Schioth, 2008).

1.4.1 Structure

P2Y family is composed of 8 subtypes that were cloned and functionally defined in human: $P2Y_1$, $P2Y_2$, $P2Y_4$, $P2Y_6$, $P2Y_{11}$, $P2Y_{12}$, $P2Y_{13}$ and $P2Y_{14}$. These receptors are phylogenetically divided into 2 subgroups, one encompassing receptors that mainly couple to G_q protein and the other to the G_i protein. The subgroup that couple to G_q protein comprised of $P2Y_1$, $P2Y_2$, $P2Y_4$, $P2Y_6$ and $P2Y_{11}$ receptors whereas the second subgroup that couple to G_i comprised $P2Y_{12}$, $P2Y_{13}$ and $P2Y_{14}$. Additionally, several other structurally related GPCR such as lipid mediators, platelet-activating factor, cysteine leukoriene and orphan receptors are clustered with both groups (Constanzi *et al*., 2004; Schoneberg *et al*., 2007). Most of the gene sequences of P2Y receptors appear to be intronless with the exception of $P2Y_{11}$. $P2Y_{11}$ receptor gene has an

intron interrupting the coding sequence after the first six amino acids (Kugelgen and Wetter, 2000).

P2Y receptors have different selectivity for adenine or uracil nucleotides, 5'-diphosphates and 5'-triphosphate. $P2Y_1$, $P2Y_{12}$ and $P2Y_{13}$ are selective for adenine nucleotides, whereas $P2Y_4$, $P2Y_6$ and $P2Y_{14}$ are selective for uracil nucleotides. $P2Y_2$ and $P2Y_{11}$ are not selective to both uracil and adenine nucleotides. Lastly, $P2Y_{14}$ has an atypical selectivity for UDP-glucose (Costanzi *et al*., 2004).

P2Y receptors share some motif of Class A GPCRs, 'DRY' and 'D/NPxxY(x)5,6F' motif at the cytoplasmic end of TM3 and TM7 respectively. Additionally, these receptors possessed signature motif of their own. All known P2Y receptor subtypes possess 4 cysteine residues at their extracellular domains which form 2 disulfide bridges: the first one between N-terminal domain and EL3 and the second bridge between EL1 and EL2 (Hoffmann *et al*., 1999). Besides that, consensus sequence 'S(I/V)(L/I)FLTCIS' is also conserved in TM3 (Kugelgen and Wetter, 2000). Although the receptors in P2Y family display a relatively high diversity in amino acid composition, conserved residues were found in TM3, 6 and 7 which are involved in ligand binding (von Kugelgen, 2005).

1.4.2 Signal transduction

$P2Y_1$, $P2Y_2$, $P2Y_4$, $P2Y_6$ and $P2Y_{11}$ receptors couple preferentially to the stimulation of phospholipase C (PLC) via G_q, and $P2Y_{12}$, $P2Y_{13}$ and $P2Y_{14}$ receptors couple to the inhibition of adenylnyl cyclase via G_i. $P2Y_{11}$ was also couple to the stimulation of adenylate cyclase via G_s. However, the specific downstream signal transduction pathway seems to be depending not only on P2Y subtype but also on cell type expressing receptor (Ralevic and Burnstock, 1998). $P2Y_{13}$ receptor couple to G_i protein subunits and subsequently inhibits adenylnyl cyclase and cAMP formation. However, transfected $P2Y_{13}$ receptor in 1321N1 neuroblastoma cells and Chinese hamster ovary can simultaneously couple to $G_{i,}$ G_{16} and also G_s upon stimulation by ADP (Marteau *et al*., 2003). Stimulation of ATP in transfected $P2Y_{11}$ receptor in 1321N1 neuroblastoma cells induced an increase in IP_3 and cytosolic Ca^{2+} as well as cAMP via G_q and G_s respectively. Stimulation with UTP unexpectedly increases only cytosolic Ca^{2+} suggesting that stimulation of ATP and UTP on $P2Y_{11}$ induces distinct signaling pathways (White *et al*., 2003).

Besides that, P2Y receptors also interact with cellular macromolecules, inducing signaling cascades that mediate responses in the cell. The interaction of P2Y receptors with small homomeric G proteins, integrins and tyrosine kinases activates mitogen-activated protein kinases (MAPK), in particularly extracellular signal-regulated kinase (ERK) pathway. In addition, P2Y receptors interact with PDZ domain and induce ion channel responses to induce trafficking in cellular membrane. The interplay of P2Y receptors with distinct G proteins and signaling pathways further indicates the complexity of this receptor family (Fischer and Krugel, 2007; Koles *et al*., 2008).

1.4.3 Biological function

P2Y receptors are involved in neuronal and non-neuronal mechanisms in regulating neurotransmission, secretion, cell division, proliferation, differentiation, regeneration and apoptosis in the cell. These receptors are broadly expressed in many tissues notably in the central and peripheral nervous system (Table 1). The wide expression of P2Y subtypes in central nervous system indicates that these receptors are part of a component to maintain physiological balance in neuronal transmission directed to the survival of the organism. P2Y receptors mediate food foraging, response to enemies as well as reaction to injury, inflammation and oxygen deficiency in the organism (Burnstock, 2006; Fischer and Krugel, 2007).

1.5 Purinergic signaling

Extracellular purines are released by the cell itself to stimulate the activation of purinergic signaling. Purinergic receptor activation subsequently trigger a host of second messenger systems and other signaling molecules including cyclic AMP (cAMP), inositol-1, 4, 5-triphosphate (IP_3), phospholipase C (PLC), arachidonic acid and nitric oxide leading to an increase of intracellular calcium concentration (Fields and Douglas, 2006).

ATP is an extracellular signaling molecule that exerts potent physiological response on a variety of tissues and cell types. ATP is released not only by presynaptic terminal but also postsynaptic membrane and other cells. The release occurs not only in response to neurotransmitter stimulation but also in response to physiological states to mechanical stress, hypoxia, inflammation and agonists. The degradation of ATP in extracellular space by ectoenzymes subsequently activates various types of P2X and P2Y receptors to provide a homeostatic regulation mechanism in the cell. P2X receptors act within milliseconds because ligand binding is directly linked to channel opening. Thus, P2X receptor can mediate fast ATP signaling over

short distances. In contrast, P2Y receptors can recognize lower concentrations of nucleotides (in the range of nM) and the binding of agonist triggers second messenger pathways. The signals and its duration is amplified and prolong over hundreds of milliseconds or even seconds, which makes it attractive as a neuromodulatory functions. Once released, ATP is rapidly degraded by ectonucleotidases to ADP, an agonist of $P2Y_{1,\ 12,\ 13}$ rceptors. The degradation of ADP in turn generates AMP and later adenosine, which stimulates postsynaptic adenosine receptors. Adenosine receptors then activates neuronal K^+ conductance and may also inhibit transmitter release via A_1 (or A_2) receptors (Fig. 1).

P2Y receptor families couple to G_q, G_s and G_i, hence modulating the activity of several types of voltage-gated ion channels via different G protein subunits. Upon stimulation with agonists, P2Y receptors inhibit N-type Ca^{2+} led to a reduced transmitter release or decreased neuronal excitability. $P2Y_{12,\ 13}$ receptors inhibit N-type Ca^{2+} channels and activate G protein-activated inward rectifier (GIRK or Kir3) K^+ channels, both via $G\beta\gamma$ subunit. $P2Y_{1,\ 2,\ 4,\ 6}$ receptors close N-type Ca^{2+} channels via G_q and inhibit GIRK and M-type K^+ channels. However, $P2Y_{1,\ 2}$ is able to open GIRK channels due to the crosstalk between G_q and G_i mediated pathways in which the receptors interact with $G\alpha_i\beta\gamma$ trimers. This interaction led to the release and binding of $G\beta\gamma$ subunits to the K^+ channel.

The release of ATP from astrocytes increased intracellular calcium via purinergic signaling ($P2X_{2,\ 4,\ 5\ ,7}$ and $P2Y_{1,\ 2,\ 4,\ 14}$). The increased cytoplasmic calcium in turn stimulates the release of glutamate and other signaling molecules (aspartate, D-serine, nitric oxide and other neuroactive substances) from astrocytes to propagate the calcium wave to adjacent cells. The interaction between neuron and glia display an intercellular communication in the brain (Fields and Burnstock, 2006; Fischer and Krugel, 2007)

1.5.1 Cell proliferation, migration and differentiation

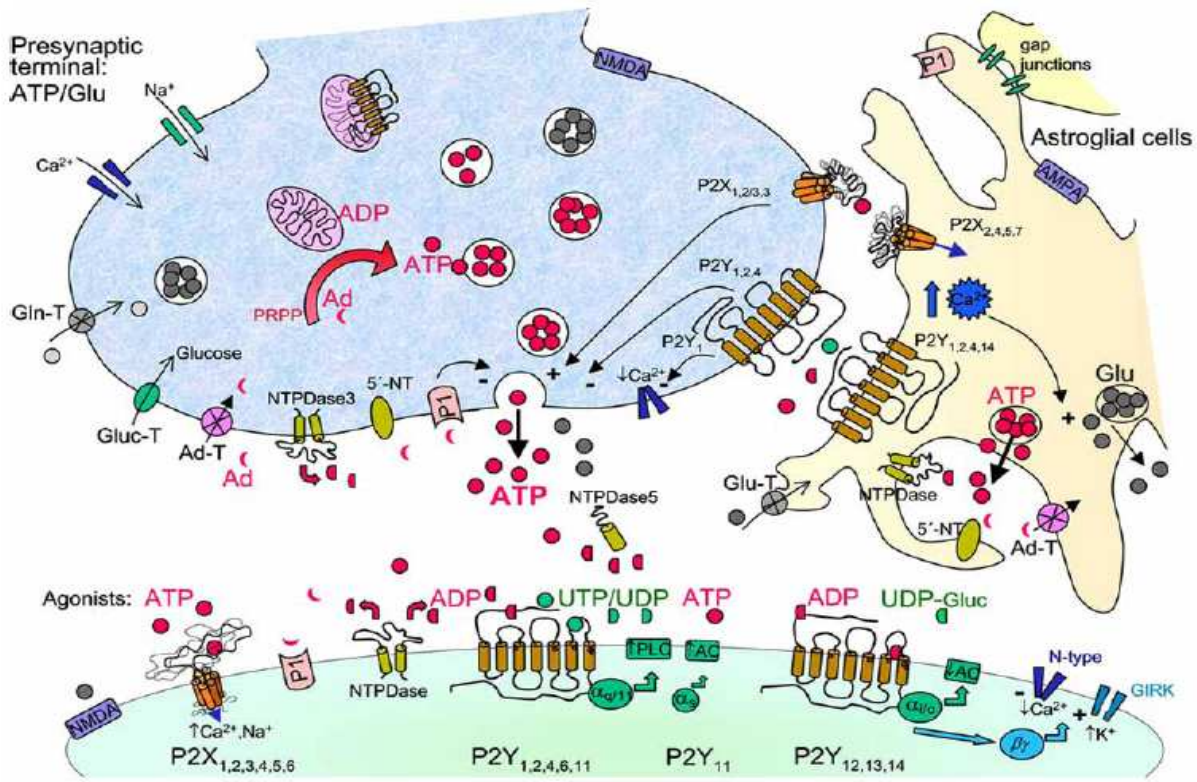
Purinergic signaling is present in the early stages of embryogenesis and is involved in processes of cell proliferation, migration and differentiation. Nucleotides exert a synergic effect on cell proliferation in association with growth factors, chemokines or cytokines in early stages of development by parallel activation of MAPK kinase pathway and/or by transactivation of growth factor receptors.

Activation of NMDA receptors increases proliferation and differentiation of neural progenitor cells (Joo *et al*., 2007). The presence of purinergic antagonists of pyridoxalphosphate-6-azophenyl-2', 4'-disulfonic acid (PPADS), reactive blue 2 or suramin in P19 neural progenitor cells impaired activity of cholinergic and glutamate NMDA receptors in differentiated P19 cells. The reduced activity of sholinergic and glutamate NMDA receptors suggest the participation of purinergic signaling in initiating downstream cascades of differentiation (Majumper *et al*., 2007, Resende *et al*., 2007).

Table 1: Characteristics of human P2Y receptors.

Receptor	Tissue distribution (Brain)	Native agonists	Principal agonists	Antagonists	Transduction
$P2Y_1$ (373 aa)	wide: placenta, prostate, heart, platelets, skeletal muscle, intestine, vascular endothelia, ***brain:*** *CC, hippoc., cerebellum(N); DRG; glial cells (astrocytes)*	ADP>ATP	MRS2365>2MeSADP >ADP≥ADPβS>>ATP (rat: 2MeSADP=2MeSATP>ADP)	MRS2500>MRS2279>MRS2179; >RB2>PPADS>suramin MRS2603 (s.$P2Y_{13}$)	G_q/G_{11} (↑PLC-β) $\downarrow I_{K(M)}$; $\downarrow(\uparrow)_{GIRK}$; $\downarrow I_{Ca(N)}$
$P2Y_2$ (377 aa)	wide: skeletal muscle, heart, spleen, bone marrow, lung, intestine, placenta, immune cells, ***brain:*** *hippoc., cerebellum (N); DRG; glial cells (astrocytes)*	UTP=ATP	MRS2698>denufosol (INS37217); UTP=ATP>Ap_4A>ATPγS (rat: UTP=ATP>CTP>GTP)	AR-C126313; tangeretin; suramin>RB2 (not PPADS)	G_q/G_{11}, (↑PLC-β) (possibly G_i/G_o) $\downarrow I_{K(M)}$; $\downarrow(\uparrow)_{GIRK}$; $\downarrow I_{Ca(N)}$
$P2Y_4$ (365 aa)	intestine, pituitary, lung, spleen, prostate, heart, lymphocytes, endothelia, ventricular system, ***brain:*** *hippoc., spinal cord (N); DRG (?); glial cells (astrocytes)*	UTP>>ATP (rat: UTP=ATP)	UTP>UTPγS>GTP>ITP (rat: UTP=ATP>Ap_4A= ITP)	PPADS>RB2; ATP (not suramin)	G_q/G_{11}, (↑PLC-β) (possibly G_i/G_o) $\downarrow I_{K(M)}$; $\downarrow_{GIRK}$; $\downarrow I_{Ca(N)}$
$P2Y_6$ (328 aa)	wide: spleen, placenta, kidney, lung, intestine, aorta, epithelia, T-cells, lymphocytes, thymus, ***brain:*** *hippoc., SN (N); DRG; glial cells (astrocytes)*	UDP>>UTP (rat: UDP>UTP>ATP)	UDPγS>UDP>UTP; INS48823 >>ADP>2MeSATP (rat: UDP>UTP>ADP>2MeSATP	MRS2578>MRS2575; RB2>PPADS>suramin	G_q/G_{11}, (↑PLC-β) (possibly G_i/G_o) $\downarrow I_{K(M)}$; $\downarrow_{GIRK}$; $\downarrow I_{Ca(N)}$
$P2Y_{11}$ (374 aa)	pituitary, spleen, immune cells; ***brain:*** *highly expr.; cell types (?)* (not apparent in rodent genome)	ATP	AR-C67085MX>ATPγS =BzATP>ATP >ADPβS>2MeSATP; UTP (?)	NF157>> 5'-AMPS; suramin>RB2 (not PPADS)	G_q/G_{11}, (↑PLC-β) and G_s (↑AC) (possibly G_o)
$P2Y_{12}$ (342 aa)	restricted: platelets, fibroblasts, endothelia, smooth muscle; ***brain:*** *astrocytes (?), oligodendrocytes, microglia*	ADP (rat: ADP>ATP)	2MeSADP>ADP>ADPβS >>MRS2365 (rat: 2MeSADP≥2MeSATP> ADP>ATP)	cangrelor (AR-C69931MX) AZD6140; AR-C78511KF*; (clopidogrel-am; prasugrel-am); suramin=RB2 (not PPADS)	G_i (↓AC) (G_o; $G_{1,2,3}$) (?) $\uparrow I_{GIRK}$; $\downarrow I_{Ca(N)}$
$P2Y_{13}$ (333 aa)	spleen, placenta, leucocytes, liver, bone marrow, lung; ***brain:*** *neuronal (?); glial cells*	ADP (rat: ADP>ATP)	2MeSADP>ADPβS>ADP (rat: ADP>2MeSADP>>ATP)	MRS2603>MRS2211>cangrelor; suramin=RB2>PPADS (not MRS2179)	G_i, (↓AC) (possibly G_{16}; G_s) $\downarrow I_{Ca(N)}$
$P2Y_{14}$ (338 aa)	wide: placenta, adipose, stomach, intestine, immune cells, lung; ***brain:*** *glial cells (astrocytes)*	UDP-glucose	UDP-glucose>UDP-galactose	not known	G_i (↓AC) (G_o/G_{16}) (?) $\downarrow I_{Ca(N)}$ (?)

Adapted from Fischer and Krugel, 2007.



Adapted from Fischer and Krugel, 2007.

Figure 1: Schematic representation of purinergic signaling at central synapses.

ATP may be produced by *de-novo* purine biosynthesis from 5'-phosphoribosyl-1-pyrophosphate (PRPP) and also by purine salvage pathways from adenosine (Ad) in neurons (and astroglial cells, not shown). ATP level in nerve terminals is generated *via* ADP mainly by oxidative phosphorylation in the mitochondria and may then be accumulated by carrier-uptake (vesicular ATP translocase) in synaptic vesicles (excitatory terminals likely present separate ATP- and glutamate-containing vesicles) as indicated. ATP released upon nerve stimulation from presynaptic terminals predominantly by exocytosis 1) activates "fast" ionotropic P2X- and with other nucleotides "slow" metabotropic P2Y-R subtypes at the postsynaptic neuronal membrane 2) modulates its own release or that of co-transmitters *via* presynaptic P2X (maybe $P2X_{1,2/3,3,7}$) and P2Y ($P2Y_{1,2,4}$) auto-receptors and 3) triggers Ca^{2+} signalling events in astrocytes *via* various P2X/Y-R (at least $P2X_{2,4,5,7}$ and $P2Y_{1,2,4,14}$). In addition, an astrocytic $[Ca^{2+}]_i$ rise can also be induced in response to neuronal glutamate release or by other neurotransmitters *via* activation of the corresponding receptors. The induced Ca^{2+} signals or propagated Ca^{2+} waves represent an important form of glial excitability and mediate glial-glial cell and glial-neuronal communication. The range of astrocyte-synapse interactions might be very complex. Subsequently, ATP and glutamate (Glu) are released from astrocytic processes enwrapping the synapse, which in turn modulate the activity of the nearby neurons and synapses or signal back to the astrocyte. Several enzyme families can degrade ATP (nucleotides). For example, the presynaptically localized ectonucleoside triphosphate diphosphohydrolase3 (NTPDase3) hydrolyzes ATP to ADP, which is a strong agonist on $P2Y_{1,12,13}$-Rs, and ADP further to AMP. The cellular localization of other NTPDase enzyme isoforms needs to be defined. Ecto-5'-nucleotidase (5'-NT), which is also present at synapses and astrocytes, degrades AMP to adenosine. Adenosine can activate various P1-Rs. Not depicted, surface-located ecto-nucleotide kinases in astroglial cells can interconvert nucleoside di- and triphosphates. Other abbreviations: Ad-T, adenosine transporter; Gln-T, glutamine transporter; Glu-T, glutamate transporter; Gluc-T, glucose transporter; $\alpha_{q/11}$, α_s, $\alpha_{i/o}$, $\beta\gamma$, different G protein signalling; (↑) indicates activation;(↓) indicates inhibition; (+) opening or (-) closure of ion channels (or positive/negative modulation of transmitter release).

1.5.2 Neuroprotection

Impaired cells release nucleic acids and their metabolites such as nucleotides, nucleosides and uric acids to the environment. These nucleic acids and their metabolites are recognized by specific host receptors (Toll-like receptors, ROG-like receptors and NOD-like receptors) and purinergic receptors (P2X, P2Y and adenosine receptors) to trigger immune responses (Ishii and Akira, 2008). Macrophages detected increased nucleotides levels via $P2Y_2$ receptor, responsive to both ATP and UTP, and initiate the release of Ca^{2+} from the endoplasmic reticulum. The calcium release, in turn, triggers surface channel to promote Ca^{2+} entry and refilling of stores (Rey *et al*., 2006). The influx of calcium is essential for T cell activation and proliferation (Carroll *et al*., 2006). Besides that, released ATP induced the activation of $P2Y_1$ in the astrocytes. Subsequently, this led to the production of interleukin-6 (IL-6) to rescue the impaired cells (Fujita *et al*., 2008).

1.6 Purine and aquatic animal

Aquatic animals can detect, discriminate and respond to a variety of water-borne stimuli to reproduce, social interactions, feeding, and predator detection (Derby and Sorensen, 2008). Likewise, fish detect an array of chemicals in their environment for habitat recognition, food finding, conspecific identification and predator avoidance. In fish, pheromone associates to habitat recognition and reproduction whereas bile acids for alarm behaviour. Amino acids and nucleotides are always associated with feeding behaviour and mediated by both olfaction and gustation.

In channel catfish, nucleotides are processed by dorsolateral of the olfactory bulb. Bile acids and amino acids are processed in the medial and lateral olfactory bulb respectively (Hansen *et al*., 2003). Likewise in zebrafish, bile salts and amino acids are processed in the medial and lateral OB respectively. Conversely, nucleotides are processed between the clusters responding to amino acids and bile salts (Li *et al*., 2005). The conservation of processing zone of odorants in olfactory bulb indicates that the odotopic map is preserved in the fish. This odotopic map is further preserved in the forebrain of channel catfish. Feeding cues of amino acids and nucleotides are represented in lateral, palllial portions of the forebrain in the channel catfish. Nucleotides responsive units were situated in the lateral half of the telencephalon near the amino acid response zone (Nikonov *et al*., 2005).

ATP is abundantly found in living organism due to their importance in carrying out biological processes. Therefore, ATP presents in high concentration in animal flesh but decays rapidly as cells die (Sikorski *et al*., 1990). Thus, ATP is an indicator for food. In spiny lobster, ATP stimulates carnivorous feeding while AMP inhibits feeding (Zimmer-Faust *et* al., 1988). In seawater, the level of ATP is minimized in the background by dephosphorylating enzymes and nucleotidases. Consequently, the presence of significant concentrations of ATP in seawater provides a reliable indicator that food source is nearby (Burnstock, 1996). The detection of extracellular ATP and other purines to initiate response in the cell is facilitated by purinergic receptor.

1.7 Purinergic receptor and olfaction

Hegg and colleagues (2003) reported that purinergic receptors modulate odor sensitivity in mice. Damaged cells release ATP, thereby activating purinergic receptors on neighboring sustentacular cells, olfactory sensory neurons, and basal cells and initiating a cascade for neuroprotection and regeneration. Calcium can modulate a number of receptors, enzymes, and channels involved in odorant signal transduction. The increase in intracellular calcium by purinergic transduction triggers homeostasis mechanisms that potentially alter odor sensitivity.

Although $P2Y_1$ and $P2Y_2$ are detected in olfactory epithelium and vomeronasal organ of the rat, both of the receptors seems not involve in chemosensation signaling. $P2Y_1$ is involved in the proliferation of epithelial cells in olfactory epithelium. Although $P2Y_2$ is detected within the neurosensory epithelium of vomeronasal organ, it was postulated that it is involved in mediating mucin secretion (Gayle and Burnstock, 2005).

In olfactory epithelium of *Xenopus tropicalis* tadpole*,* ATP activates a large number of sustentacular cells and to a lower extent, olfactory sensory neurons. Upon activation by ATP, sustentacular cells may secrete components that influence the composition of the mucus in olfactory epithelium and thereby, could affect the olfactory perception (Czesnik *et al*., 2006).

P2Y receptor could be olfactory receptor or could be the charaperone of OR. The association of purinergic receptors ($P2Y_1$, $P2Y_2$ and A_{2a}) with olfactory receptor M71 in HEK-293 cells enhances plasma localization of M71. The enhanced plasma localization with purinergic receptors was also observed in M71 subfamily members, but not distantly related members. This association suggesting that the interaction purinergic receptor is confined to a restricted

population of OR. The OR interactions with other receptors may act in concert with OR associations with accessory proteins to control OR trafficking (Bush *et al*., 2007).

1.8 Zebrafish as model organism

The zebrafish is an important model system for genetic and developmental studies (Kimmel, 1993). It has the combination of the best virtues of all other models which includes easily accessible embryos, simple breeding and short generation time. Besides that, the embryos are completely transparent during early development facilitating visualization. The organs in zebrafish larvae composed of fewer cells compared to other vertebrate model even though the organ functions in much the same way as in other vertebrates. All these attributes make zebrafish a desirable model to work with.

1.9 Aims:

The aim of this thesis is to lay the foundation for a comprehensive analysis of the P2Y receptor family in aquatic model system, zebrafish. Previously, most of the P2Y receptor studies are conducted in pharmacology research. Only a few studies have been conducted in vertebrate development. In particular, nearly nothing is known about the possible role of P2Y receptors in zebrafish, Hence, the aims of this project are:

i) To identify P2Y receptors and delineate the family in teleost
ii) To understand the evolutionary dynamics of P2Y receptor
iii) To characterise expression of P2Y receptor genes

2. RESULTS

2.1 P2Y-like genes in teleost

A recursive database search against 5 teleost species (*Danio rerio*, *Gasterosteus acualeatus*, *Oryzias latipes*, *Tetraodon nigroviridis* and *Takifugu rubripes*) with eight functional human P2Y receptors as a query has led to the identification of 42 orthologues. Further bioinformatic analysis revealed an additional 185 GPCRs that showed a high degree of structural similarity with P2Y receptors. (Zebrafish (*Danio rerio*) have the largest family, with 68 genes. This is followed by stickleback (*Gasterosteus acualeatus*) and medaka (*Oryzias latipes*) whose P2Y families are comprised of 48 and 40 genes respectively. Lastly, 35 genes and 36 genes constitute the tetraodon (*Tetraodon nigroviridis*) and fugu (*Takifugu rubripes*) P2Y families, respectively.

To analyse the genesis of the P2Y family, database searches were performed in jawless and cartilaginous fishes. Twenty four P2Y genes were identified in the lamprey (*Petromyzon marinus*) genome, and 24 genes in the elephant shark (*Callorhinchus milii*) genome. The number of genes in these two species might not be the complete number as the genome projects are still in progress, and the genome coverage for shark is only 1.4 fold. A single *p2y*-like gene was found in little skate (*Raja erinacea*), another cartilaginous fish with only partial genomic sequence information available (Table 2).

To pinpoint the evolutionary origin of the P2Y family, database searches were extended to the ancestral chordate species, *Ciona intestinalis*, an ascidian. Only one candidate with rather low homology was found, but it turned out not to fulfil the inclusion criteria (see Methods), as it did not exhibit the seven transmembrane domain structure.

Intriguingly, the number of P2Y-like genes is not much different between *Xenopus tropicalis* and human (39 and 33 genes respectively). This suggests that those genes retained in the vertebrate transition from water to land may have kept their function. In contrast, large differences in family size are observed between teleost species, suggesting a dynamic adaptation to changing environmental conditions.

Table 2: P2Y genes in 10 vertebrate species.

	Dr	Ga	Ol	Tn	Tr	Re	Cm	Pm	Xt	Hs
Total	68	48	40	35	36	1	24	24	39	33
Funtional genes	67	45	38	31	33	1	19	20	36	33
Pseudogenes or partial sequences	1	3	2	4	3	-	5	4	3	-

Abbreviations: Dr, zebrafish; Ga, stickleback; Ol, medaka; Tn, tetraodon; Tr, fugu; Re, little skate; Cm, elephant shark; Pm, lamprey; Xt, frog; Hs, human.

2.2 P2Y genes are paraphyletic

To investigate the phylogenetic relationship of the P2Y receptors and candidate proteins obtained above, methods of neighbour joining (NJ), maximum parsimony (MP) and maximum likelihood (ML) are used. The three methods gave slightly varying tree topologies with the ML tree having the best resolution (Fig. 2). Olfactory receptors were used as an outgroup. Both olfactory receptor and P2Y receptor belong to the delta subgroup of GPCR class A receptor family. Hence, the olfactory receptor family is the nearest phylogenetic relative of the P2Y receptor family. Phylogenomic analysis revealed that P2Y receptor family is paraphyletic. Within the P2Y clade two other gene groups are contained, platelet-activating factor receptors and succinate receptors. These 'internal outgroups' are interspersed between two phylogenetically distinct subgroups of P2Y receptors. This is consistent with an earlier report by Abbracchio *et al* (2006). Within the P2Y clade a large number of orphan GPCRs are clustered together with the functional P2Y receptors, suggesting the presence of yet to be characterized P2Ys. Moreover, leukotriene cysteine, EBI2 and free-fatty acids receptors are close neighbors in the phylogenetic tree indicating that P2Y and related receptors exhibit high diversity both in sequences and ligands.

According to the International Union of Pharmacology, the term 'P2Y' is used for functional receptor proteins and the lowercase 'p2y' is used for mammalian orphan receptors or functional nonmammalian receptors without a mammalian orthologue. Hence, I have named the fish sequences according to their human orthologue. In the case that non-nucleotide ligands are known for a mammalian receptor, the orthologous fish sequence is named after the mammalian gene. The receptors within the P2Y family with unknown ligand are named as p2y-like (p2y-l) as they share a high sequence homology to P2Y receptor (Fig. 3).

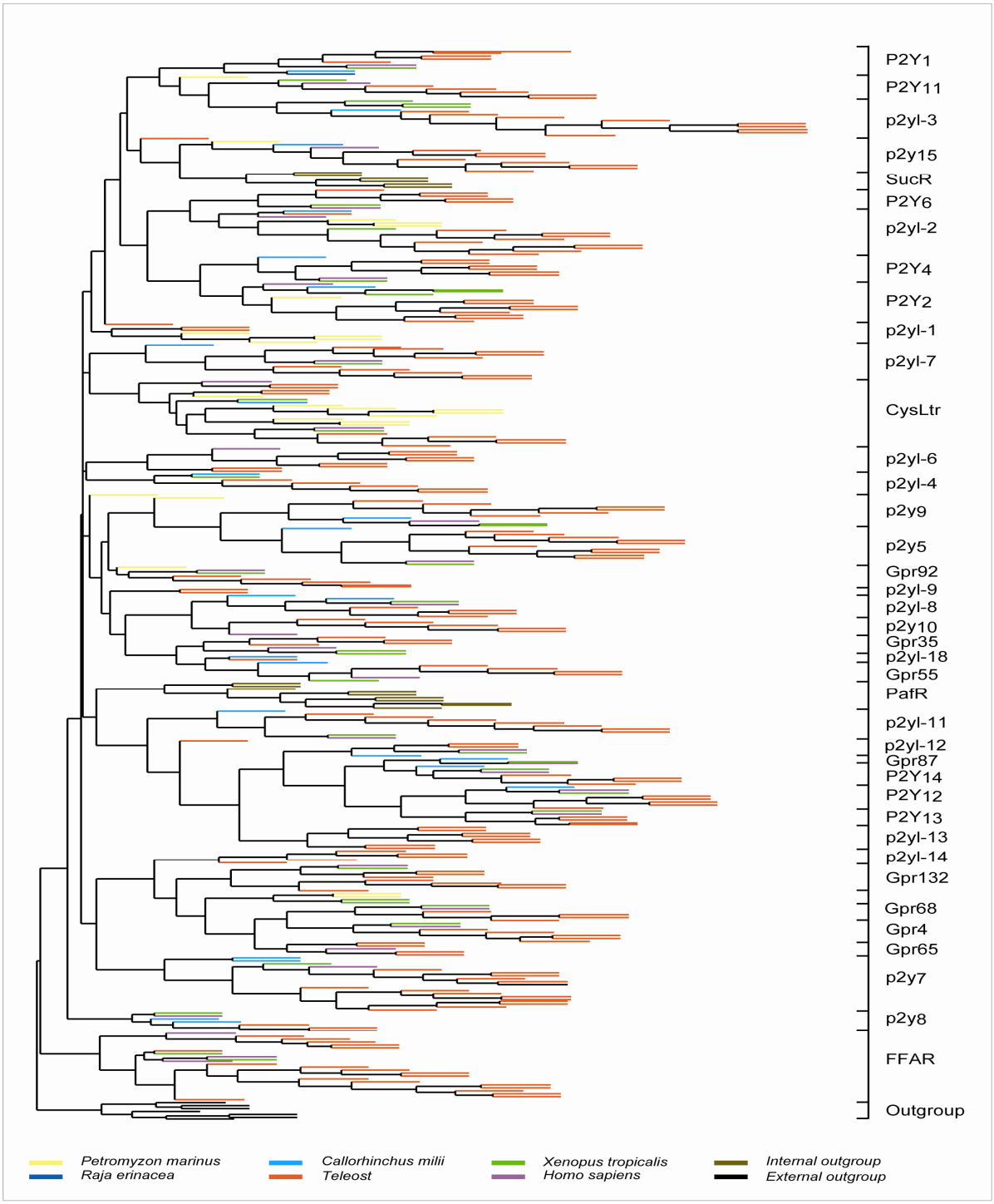


Figure 2 | Phylogeny of teleost with other vertebrate P2Y families.

A maximum likelihood tree was constructed based on the alignment of predicted amino acid sequences with 100 bootstrap replicas. 227 teleost P2Y (red), 24 shark P2Y (light blue), 1 little skate (dark blue), 24 lamprey P2Y (yellow), 28 frog P2Y (green) with human P2Y (purple). Olfactory receptors (black) were used to root the tree. Both platelet-activating factor and succinate receptors served as internal outgroups of the P2Y family and are labeled in brown.

2.3 P2Y gene gains and losses in the teleost and tetrapod lineage

All P2Y genes identified clustered into 19 major clades divided into 38 subfamilies. Nearly all of the subfamilies contain tetrapod and teleost genes in each clade demonstrating that members of the subfamily have established before the divergence of *Actinopterygii* and *Sarcopterygii*. However, two of the subfamilies did not have a mammalian orthologue (p2yl-3 and p2yl-4) even though the orthologue is present in *Gallus* and *Xenopus* (Appx 7.2A). The absence of the mammalian orthologue implied that the vertebrate orthologue is lost in the mammalian lineage. On the other hand, several gene expansion events occurred in the teleost fish lineage, in eight of the subfamilies ($P2Y_2$, $P2Y_4$, $p2y_5$, $p2y_7$, $p2y_{15}$, p2yl-7, p2yl-13 and p2yl-14). These gene duplications have occurred at different stages in evolution, as visualized by their presence in different subsets of teleost species. This scenario is displayed by $P2Y_4$ and $p2y_{15}$ clades in which the expansion only occurred in zebrafish, whereas $P2Y_4$ and $p2y_{15}$ remained a single gene in other teleosts. Furthermore, species-specific expansion in p2yl-13 clade only involved zebrafish (Fig. 3) and stickleback.

More than 60% of the clades have the full set of orthologues for each of the five analysed teleost fish genomes. It cannot be excluded that the lack of some orthologues in the remaining clades might be due to the inadequacy of the currently available databases. However, there is also the possibility that these genes are lost in some species.

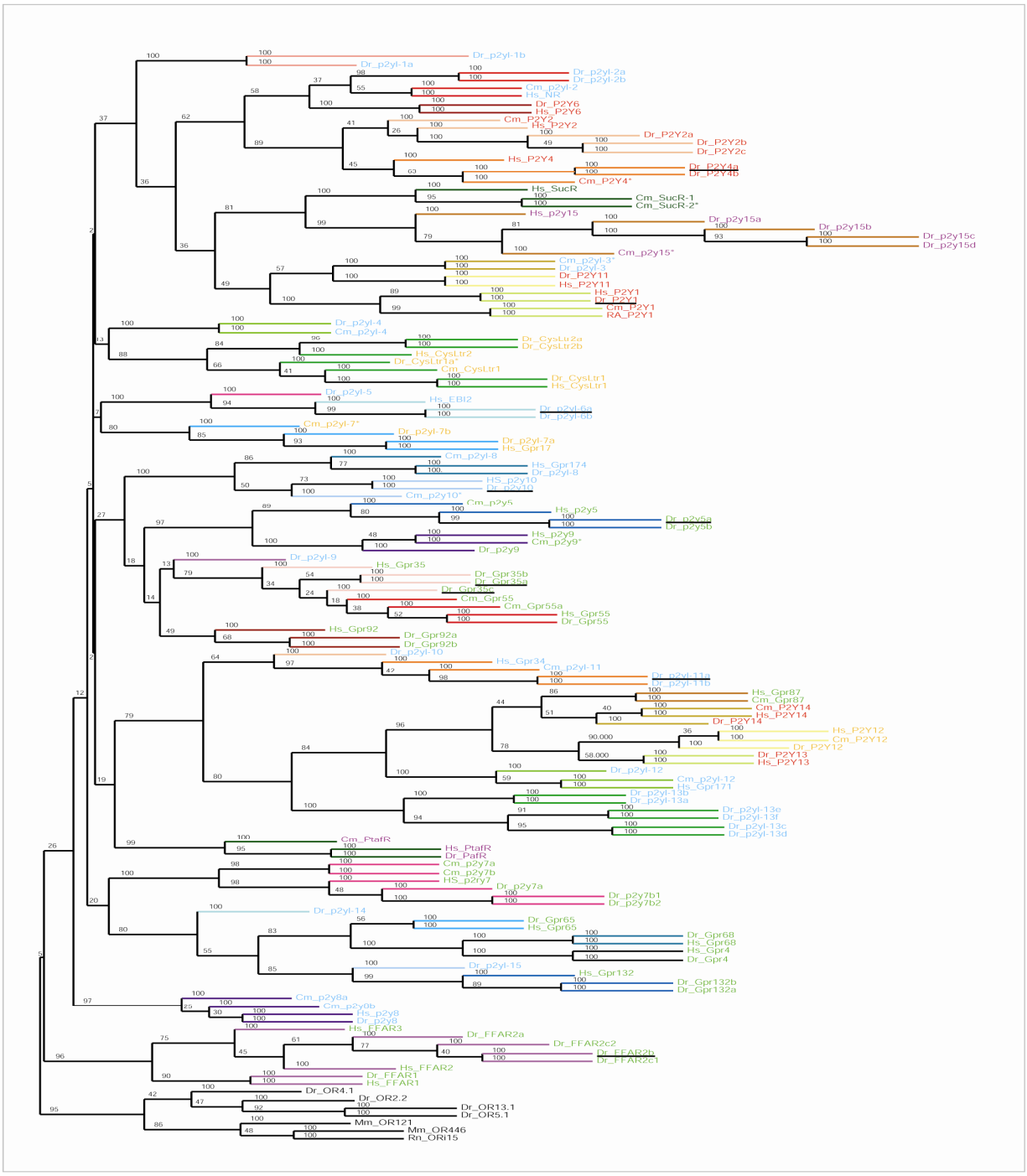


Figure 3 | Phylogeny of zebrafish P2Y gene and ligand correlation.

Phylogenetic tree of P2Y genes identified from zebrafish (68 genes), shark (24 genes) and human (33 genes). The tree was constructed using ML method with 100 bootstrap replicas. The branches are coloured to facilitate the identification of subfamilies. Dark green and black braches indicate internal and external outgroups respectively. The gene name is colour-coded according to the ligand: red, nucleotides; blue, lipid derivatives; orange, nucleotides and lipid derivatives; purple, α-ketoglutarate; green, unknown ligand. Underline gene names are candidates used for subsequent analysis.

2.3 Correlation between phylogenetic location and ligand

To examine the relationship between phylogenetic location and known ligand specificity, I have graphed the ligand specificity onto the maximum likelihood tree in Figure 3. Ligand information was collected for human P2Y receptors. Most receptors in this family bind either to nucleotide or lipid derivatives. Phylogenetic clusters of receptors that bind to similar types of ligands are present. However, this correlation is only partial, and there are several examples of phylogenetically neighboring receptors with rather different ligands, and also examples of phylogenetically distant receptors sharing a ligand.

Some examples of ligand promiscuity of the receptor are the activation of $P2Y_{12}$ by both ADP and leukotrienes (Nonaka *et al*., 2005) and the activation of cysteine leukotriene receptor 1, cysteine leukotriene receptor 2 and Gpr17 by leukotrienes as well as UDP at nanomolar concentration (von Kuegelren, 2005). The large difference in chemical structure of these two endogenous ligands indicates the complexity of the ligand-receptor interactions in the diverse P2Y family.

The primitive P2Y-receptor found in the little skate *Raja erinacea* does not discriminate between different nucleotides agonists such as ATP, ADP or UDP (Dranoff *et al*., 2000). In contrast, mammalian $P2Y_1$ which forms a sister clade with *Raja erinacea* P2Y-receptor is potently activated only by adenine nucleotides and not by uracil nucleotides (Fischer and Kruegel, 2007). Thus, ligand selectivity of orthologues exhibits major species differences indicating a fine-tuning process during evolution. Hence, ligand-binding analysis of the P2Y repertoire with a whole range of species will be instructive.

2.4 P2Y genes have an early origin in jawless fish

The presence of P2Y orthologues in jawless and cartilaginous fishes revealed an ancient origin of this family of receptors. Sea lamprey P2Y genes usually form very basal branches (Fig. 2) inferring that the ancestral genes for these families existed before the segregation in jawless and jawed fish, about 560 million years (Volff, 2005). The phylogenetic analysis displayed that sea lamprey receptors did not always have one-to-one orthologue with human and teleost species. In some cases, sea lamprey representation is missing in the clade. There are two possibility of the absent in the clade, either the ancient sea lamprey gene have been lost during the transition from jawless to jawed fish or inadequacies of the current sea lamprey genome database.

The clustering of sea lamprey genes, basal to the three subfamilies (p2yl-1, p2yl-2, cysteine leukotriene receptor) indicating an independent duplication event occurred in lamprey lineage. These genes formed a paralogue group within lamprey lineage. In few cases, sea lamprey gene is the ancestral of duplicate genes of higher species. Sea lamprey gene, P2Y*u* is ancestral gene of p2y5 and p2y9. The duplication of this orthologue occurred after the divergence from the jawless vertebrate.

On the other hand, elephant shark P2Y genes are distributed uniformly in the tree. Almost all the clades have orthologues shark gene and phylogenetically, the gene is located deep within the teleost clade demonstrating that the P2Y families are already established in the common ancestor of shark and bony fish.

Surprisingly, lamprey and elephant shark orthologues are not present in free-fatty acid receptor clade. The absence of the representatives from basal vertebrate raises the question whether free–fatty acid receptors arise only after the divergence of bony vertebrates. This will remain to be elucidated until both the lamprey and elephant shark genomes have been fully sequenced.

The distribution of the P2Y gene in the phylogenetic tree closely reflects the phylogenetic relationships of the species. The two pufferfish studied belong to the same family (*Tetraodontidae*) and their orthologues are closer related to each other than to those from any other species. The stickleback and medaka orthologues constituted their next neighbors, as pufferfish, stickleback and medaka belong to three subdivisions of the same superorder *Acanthopterygii* (orders *Tetraodontiformes*, *Gasterosteiformes*, and *Beloniformes* respectively). In all cases the zebrafish orthologues occupy the most distant position (zebrafish being the evolutionarily most distant fish in this comparison, as it is an *Ostariophysi*, unlike the four other species which belong to a *Neoteleostei*.

2.5 Global negative selection in teleost P2Y genes with a few positives sites

To investigate the evolutionary dynamics of P2Y genes, selective pressure analysis is conducted in teleost P2Y genes using both global and local analysis of substitution rates in non-synonymous (dN) vs. synonymous (dS) sites in 31 of the subfamilies. The remaining seven subfamilies are not included in the analysis due to the lack of number of genes. The global dN/dS values calculated for each of the groups ranging from 0.08 to 0.42, with an average of 0.23, indicating that P2Y genes are subjected to purifying selection (Fig. 4a). The purifying

selection is against amino acid changes in protein-coding genes, hence suggesting that these genes are slow evolving genes.

Global analysis of negative selection could mask positive selection on a few amino acid sites. Site-by-site dN/dS analysis of the P2Y genes revealed a positively selected site in two of the analysed groups, p2yl-4 (Fig. 4d) and p2yl-7 (Fig. 4e). These sites are located in TM5 and N-terminal in p2yl-4 and p2yl-7 respectively. Conserved residues in TMs 3, 6 and 7 are likely to contribute to the binding pocket of the P2Y receptor (Erb *et al*., 1995; Jiang *et al*., 1997; Moro *et al*., 1998; Hoffman *et al*., 1999; Moro *et al*., 1999; Qi *et al*., 2001; Guo *et al*., 2002; Harold *et al*., 2004). All the positively selected sites are not positioned at the binding pocket indicating that the ligand specificity is most likely conserved in teleost lineage.

2.6 High conservation in transmembrane domains

To analyse the sequence similarity of the P2Y genes, pairwise comparisons are performed. P2Y genes showed low identity values (20-50%) indicating that this class of receptors is highly divergent in the amino acid composition (Appx. 7.2C). Across the species, homologues of a subtype have a higher homology and sequence conservation. For example, in the $P2Y_1$ clade, zebrafish shared more than 60% of identity with little skate, elephant shark and human. The shared identity is extended to more than 80% in teleost species (data not shown).

P2Y receptor belongs to the rhodopsin family of GPCR. Therefore, distinct amino conservation is much more prevalent in transmembrane region than non-transmembrane region. Nine sequence motifs of rhodopsin family is conserved in P2Y receptors (‘N’ in TM1, ‘L’ in TM2, ‘C’ in EC1 and EC2, ‘R’ in IC2, ‘W’ in TM4 and, ‘N’, ‘P’ and ‘Y’ in TM7). Besides that, P2Y also shared a conserved amino acid with VIR and ORA respectively (‘G’ in TM1 and ‘S’ in TM3). In addition, a conserved proline (P) in TM5 is also present in ORA, T2R and TAAR suggesting that this is a general motif of rhodopsin family (Fig. 5).

Next, consensus motifs are analyzed to search for characteristic sequence motif of this receptor family in fish. There are three consensus motifs reported for functional P2Y receptors in mammals. First, all known P2Y receptor subtypes possess 4 cysteine residues at their extracellular domain (Hoffmann *et al.,* 1999; Ding *et al.,* 2003). Second, a conserved consensus sequence of “S(I/V)(L/I)FLTCIS” in TM3 (Kuegelgen and Wetter, 2000). And third, an “Hxx(R/K)” motif is present in TM6 (Abbracchio *et al.,* 2006). The internal outgroups, platelet-activating

factor and succinate receptors possessed two of the three motifs listed above, emphasizing the high structural similarity with P2Y receptors. Both platelet-activating factor and succinate receptors possessed "S(I/V)LFLxxIS, HxxQ" and "SILF(L/M)xxxS, HxxR" motifs at TM3 and TM6 (data not shown). The conservation of cycteine residue is also extended to the outgroups. Platelet-activating and succinate receptors retained cysteine residue at two and three of the extracellular domain respectively. However, conserved cysteine in N-terminal is absent in both receptors. Therefore, the cysteine residue in N-terminal may serve as a signature motif in identifying members of P2Y family.

The analysed subfamilies that possesed 4 cysteine residues in the extracellular loop are grouped into 3 classes: i) P2Y receptor; ii) uncategorized receptors that have known ligand (gpr55, gpr65, gpr68, gpr132 and cysteine leukotriene receptors) and iii) orphan receptors that have no known ligand (p2yl-2, p2yl-3, p2yl-4, p2yl-6, p2yl-7, p2yl-8, p2yl-11, p2yl-12, p2yl-13 and p2yl-14). This motif shared by P2Y receptor and orphan receptors may help to identified the yet remaining P2Y members.

Although the receptors possessed conserved sequence motifs, most of the motifs are also present in platelet-activating and succinate receptors. Both the receptors intertwined P2Y receptors in the phylogenetic tree. Hence, it is not surprising for platelet-activating and succinate receptors to share a high sequence similarity to P2Y receptor. Despite the high conservation of P2Y receptor with other receptors of rhodopsin family, yet three P2Y-specific motifs are located; 'C' in N-terminus, 'G' in EC1 and 'Y' in TM5. All the 3 P2Y-specific motifs are also present in the mammalian P2Ys.

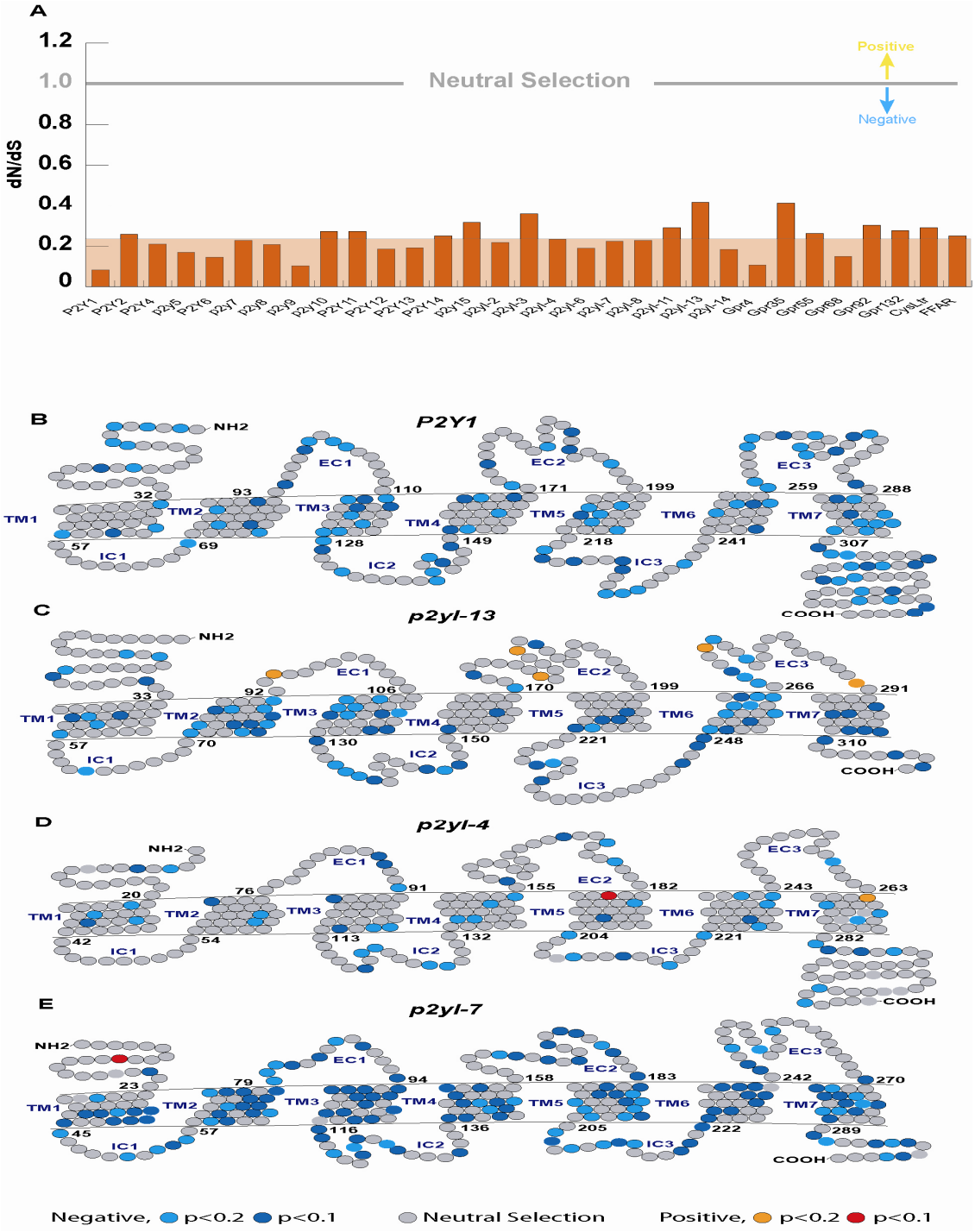


Figure 4 | Evolutionary distances and selective pressure on P2Y genes.

(A) dN/dS ratios of 31 teleost P2Y subfamilies. dN/dS analysis is conducted only in group that contain more than 2 genes. dN/d >1 indicates positive selection; dN/dS < 1, purifying selection; and dN/dS =1, neural selection. The group average ratio is indicated by background shading. (B-E) A representative site-by-site selective pressure is shown for four P2Y sequences (negative selection in light blue, p<0.2 or blue, p<0.1, neutral selection in grey, positive selection in orange, p<0.2 and red, p<0.1.

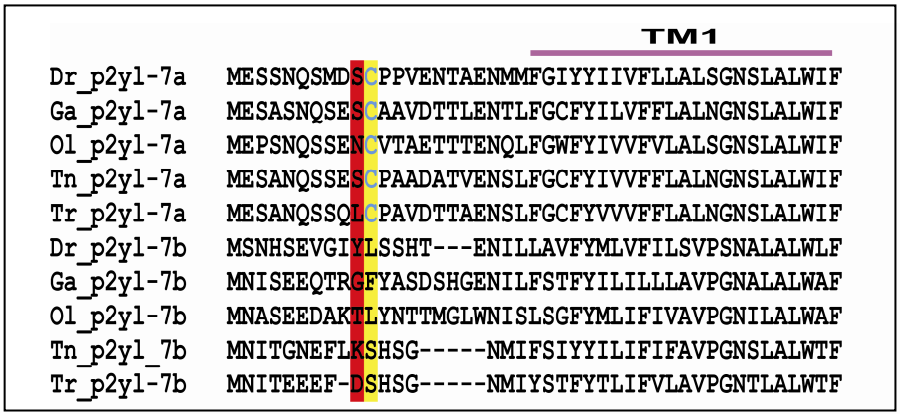


Figure 5 | Positively selected amino acid site in teleost p2yl-7.

Red shaded indicates positively selected amino acid site in N-terminal. Yellow shaded indicates conserved cysteine residue site in N-terminal. Conserved cysteine is labeled in blue.

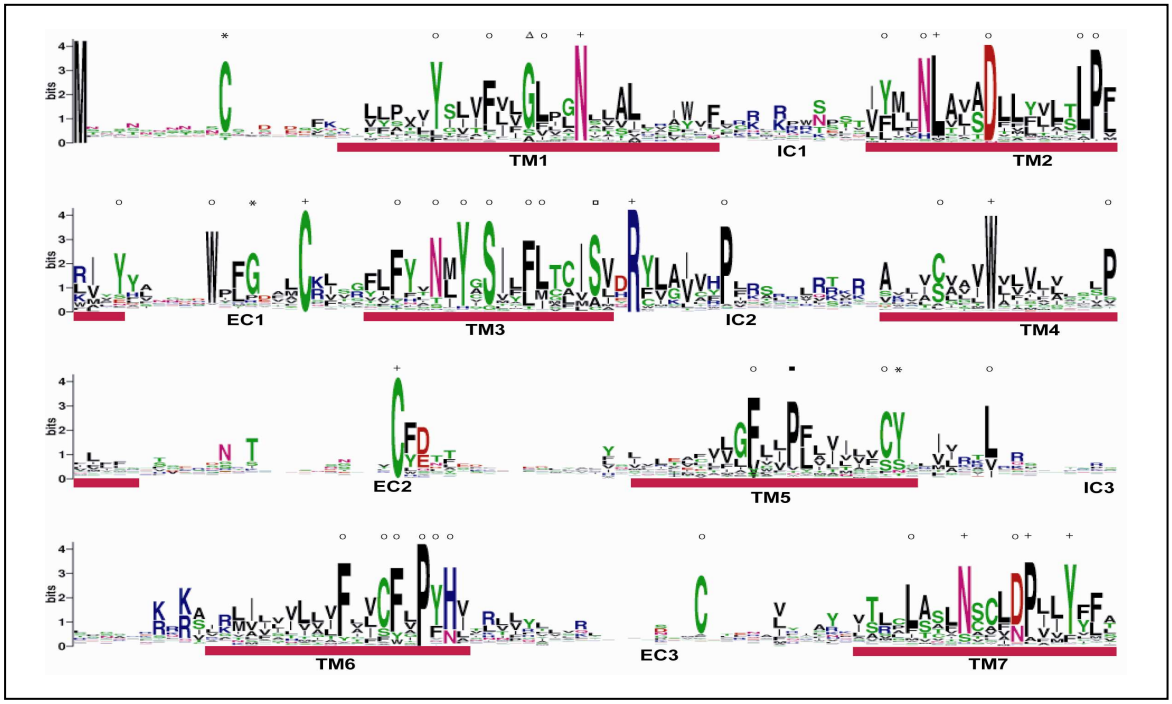


Figure 6 | Conserved sequence motifs of P2Y family.

Conservation of predicted amino acid sequence for the fish P2Y repertoire is displayed as sequence logo. The regions corresponding to the transmembrane (TM) domains and the extracellular and intracellular (EC and IC) domains are numbered and indicated. Crosses represent residues generally conserved among other class A GPCR families and circles represent residues conserved in platlet-activating and succinate receptors. Triangles represent residues conserved between P2Y and V1R. Squares represent aresidues conserved between P2Y and ORA. Filled squares represent residues conserved between P2Y, ORA, T2R and TAARs. Asterisks represent conserved residues found only in P2Y.

2.7 Genomic distribution of zebrafish P2Y genes

In zebrafish genomes, P2Y genes are distributed broadly, residing on 20 chromosomes. The genes are distributed both as singleton and cluster (Fig. 7). There is one major cluster containing 10 genes on chromosome 15. Several small clusters comprise of two to four genes are also found on chromosome 1, 2, 5, 7, 9, 14, 17, 18 and 19. The rest of the genes existed as singletons residing in other chromosomes except on chromosome 8, 11, 12, 22 and 25. Only 5 of the identified P2Y genes remain unassigned to any chromosomes.

The genomic location of ten listed genes in the major cluster in chromosome 15 concurred with the phylogenetic analysis inferring the presence of expansion event. Furthermore, the genes in the cluster are arranged in the tandem as in phylogenetic analysis postulating the existence of correlation between genomic linkage and phylogenetic relationship.

The ten genes are arranged in the order of: $P2Y_{12}$, $P2Y_{13}$, p2yl-13b, p2yl-13a, $P2Y_{14}$, p2yl-13e, p2yl-13f, p2yl-13d, p2yl-13c and p2yl-12 (Fig. 8a). The cluster composed of $P2Y_{12}$, $P2Y_{13}$, $P2Y_{14}$ and p2yl-12 is evolutionary well-preserved. This is supported by shared genomic location in elephant shark (cluster 405), stickleback (chromosome I), medaka (chromosome 13), tetraodon (chromosome 16) and human (3q24-3q25) (Schoeneberg *et al.*, 2007; Venkatesh *et al.*, 2007). The addition of p2yl-13 subfamily into gene cluster of zebrafish demonstrated the gene expansion via gene duplication.

Two intrachromosomal duplication events have taken place. The first duplication event generated p2yl-13a and p2yl-13b from p2yl-12. Then, it was followed by second duplication event, producing p2yl-13c, p2yl-13d, p2yl-13e and p2yl-13f on chromosome 15 in zebrafish. The four genes may be originated from p2yl-13a and p2yl-13b based on the phylogenetic relationship. Besides zebrafish, the second duplication occurred only in stickleback although the duplicated gene is translocated to a different chromosome.

Second duplication event is not presence in other fish species suggesting either the duplication event is limited to zebrafish and stickleback only or the cluster of genes is lost in other species during evolution process.

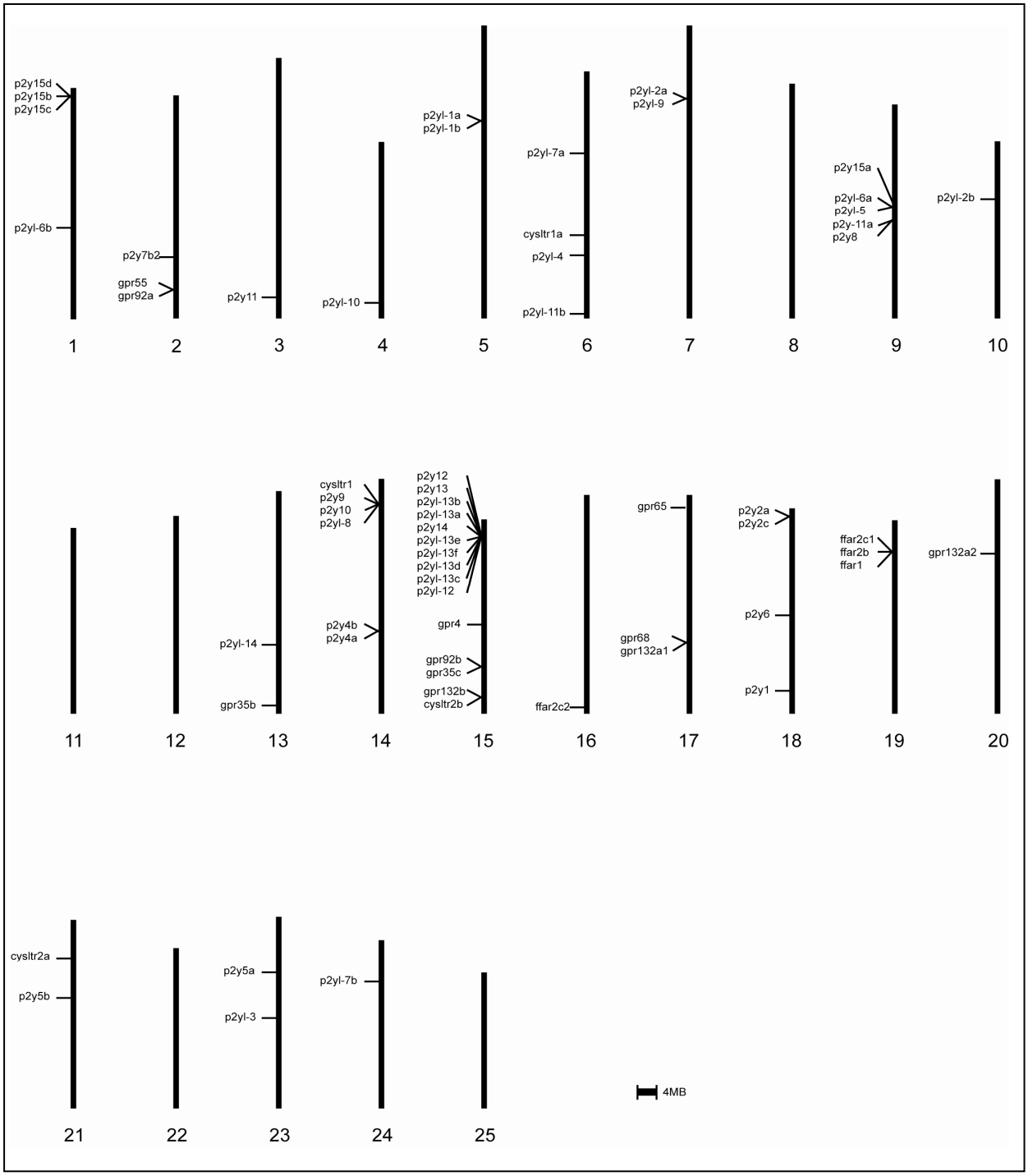


Figure 7 | Chormosomal distribution of zebrafish P2Y genes.

The P2Y genes are distributed in both as singletons and small clusters in the zebrafish genome. The genes are distributed in 20 chromosomes and only 5 genes remained unassigned in any of the chromosome.

2.8 Syntenic relationships of teleost P2Y

The comparison of genomic location of P2Y genes in zebrafish, stickleback, medaka and tetraodon revealed a remarkable degree of synteny across the fish species. Two homologous segments are identified. The identified segments possessed synteny distances similar to phylogenetic relationship.

The first paralogon has a large number of genes consist of both P2Y receptors ($P2Y_1$, $P2Y_2$, $p2y_5$, $P2Y_6$, $P2Y_{12}$, $P2Y_{13}$ and $P2Y_{14}$) and orphan receptors (p2yl-11 p2yl-12, p2yl-13, Gpr4, Gpr92, Gpr132). The segment displayed disrupted synteny as the genes are scattered across 3 chromosomes in zebrafish and tetraodon, 2 chromosomes in medaka and 1 chromosome in stickleback (Fig. 8a). The second paralogon displayed conserved synteny. The genes ($P2Y_4$, $p2y_9$, $p2y_{10}$, p2yl-8 and cycteine leukotriene receptor 1) included in the second homologous segment are distributed in a single chromosome in all analysed species (Fig. 8b). Stickleback and tetraodon shared a high conserved linkage. Both species shared the conservation of both in synteny and order of homologous genes.

In both segments, members of syntenic groups of medaka are more susceptible to inverted order. The inverted order of syntenic groups implying a higher frequency of intrachromosomal rearrangements event in medaka genome..

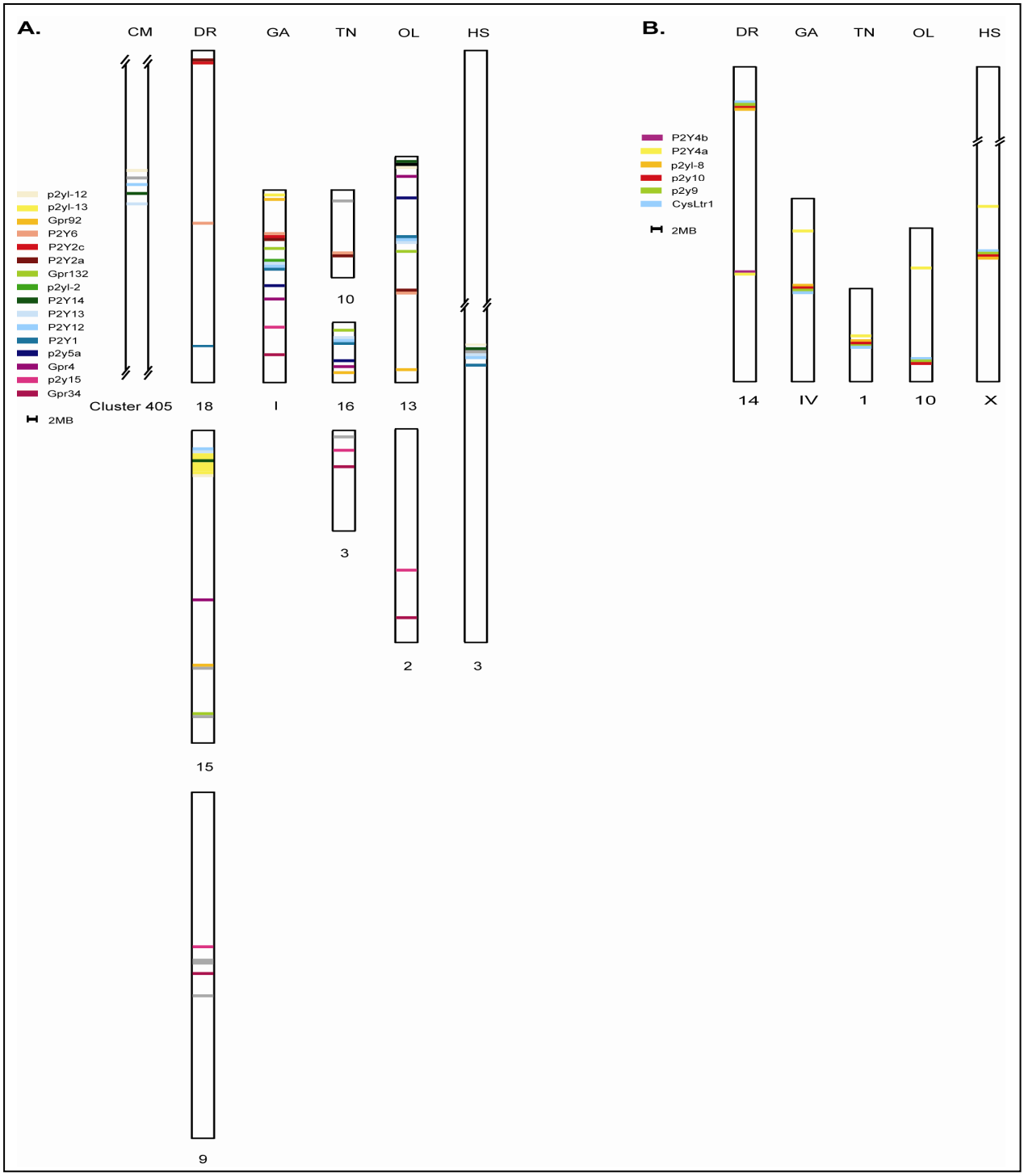


Figure 8 | Conserved synteny in fish species and human.

Conservation of synteny of genes in zebrafish chromosomes compared to *Gasterosteus aculeatus* (GA), *Tetraodon nigroviridis* (TN) and *Oryzias latipes* (OL). Genes are ordered according to their genomic positions. Grey bar indicates non-conserved gene. Synteny of genes residing in chromosome 9, 15 and 16 (A) and, chromosome 14 (B) in zebrafish compared to 3 other fish species. Grey bar represents other genes.

2.9 Genomic structure of P2Y genes

Currently, human P2Y receptors are the well-studied in comparison to other available species. Most of the human functional P2Y receptors do not contain introns. Only one of the receptor, P2Y11, has an intron after the third amino acid. In contrast, the analysed genes of seven fish species and a frog species possessed a mixture of gene structure, varied from intronless to multiexonic structure (Fig. 9). Most of the analysed genes (75%) are intronless. The remaining 25% of the analysed genes possessed intron, with the number of introns varied from one to five (Table 3). At present, all the analysed genes in lamprey and elephant shark are intronless. However, the intronless structure is not well-preserved in bony fish. This is displayed by the presence of multiexonic structure suggesting possibly an intron-gain event occur in the bony fish evolution.

In zebrafish, 81% of the genes are intronless, 13% acquired an intron, 4% acquired 2 introns and only 2% acquired 3 introns in their gene structure. On the contrary, the percentage of intronless gene in neoteleost species is much reduced ranging from 56% to 63%. The proportion of genes that possessed multiexonic structure is increased. Additionally, the number of insertion of introns extended to 5 introns in stickleback, tetaodon and fugu (ranging between 2 to 3%).

The multiexonic genes are scattered across the phylogenetic tree suggesting that the intron-gained event is a random process. Frequently, if a member of the clade acquired an intron, most of the remaining members would acquire introns as well (e.g $P2Y_{13}$). Generally, neoteleost species would acquire an additional insertion of intron in comparison to zebrafish, an ostariophysan. Occasionally, only one member of the clade acquired intron whereas other members of the clade remained intronless as in the case $p2y_{5b}$ in medaka, in which an intron-gain event took place. A very rare and interesting case is in $p2y_{10}$ clade. Every member of the clade acquired a different number of intron. Zebrafish acquired an intron in the gene structure whereas stickleback, medaka and fugu acquired two, three and four introns respectively. Tetraodon $p2y_{10}$ sequence is incomplete and currently, remained intronless. This variance displayed the dynamicity of intron-gain event in different genes and lineages. The intronless structure of P2Y gene in lamprey is well-preserved in the mammalians, given that almost all tetrapod P2Y genes appear to be monoexonic. In contrast, teleost fishes appear to be susceptible of insertion of intron. The differences in gene structure preservation between mammals and teleost which share a common ancestor suggests that a higher propensity of intron insertion in *Actinopterygii* lineage but less in *Sarcopterygii* (Fig. 10).

Table 3: Gene structure distribution in 8 vertebrate species

Organism	Intronless	1-intron	2-intron	3-intron	4-intron	5-intron	Total
Danio rerio	55 (81)	9 (13)	3 (4)	1 (2)	-	-	**68**
Gasterosteus aculeatus	27 (56)	9 (19)	8 (17)	3 (6)	-	1 (2)	**48**
Oryzias latipes	24 (60)	9 (22)	4 (10)	3 (8)	-	-	**40**
Tetraodon nigroviridis	22 (63)	7 (20)	4 (11)	1 (3)	-	1 (3)	**35**
Takifugu rubripes	22 (61)	7 (19)	3 (8)	2 (6)	1 (3)	1 (3)	**36**
Callorhinchus milii	24 (100)	-	-	-	-	-	**24**
Petromyzon marinus	24 (100)	-	-	-	-	-	**24**
Xenopus tropicalis	38 (97)	1 (3)	-	-	-	-	**39**
Total	**236 (75)**	**42 (13)**	**22 (7)**	**10 (3)**	**1 (1)**	**3 (1)**	**314**

The genes in each species (row) are grouped according to their structure, varied from intronless to multiexonic (column). The number indication of total of genes in each species that possessed the specified structure and the percentage is listed in parenthesis.

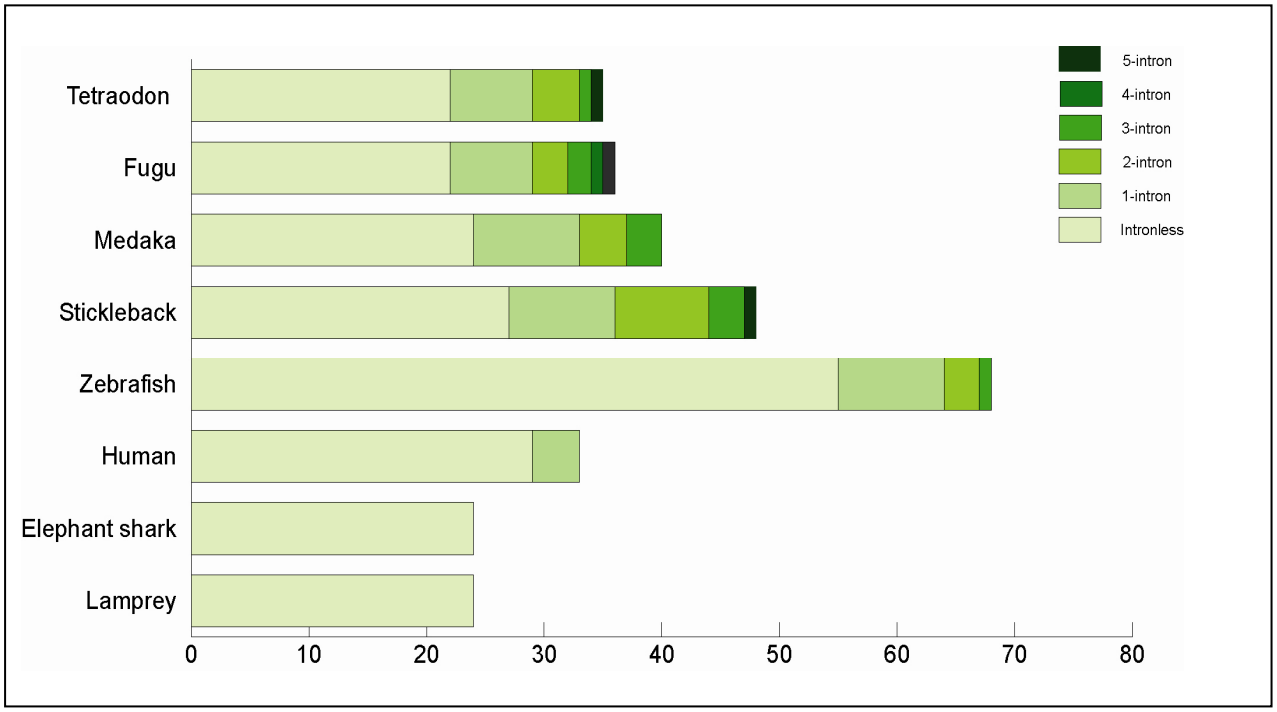


Figure 9 | Gene structure distribution in 7 fish species and human.

The P2Y genes have a mixture of intronless and multiexonic structure in fish species and human. The distribution of the gene structure is depicted in stacked bar graph. X-axis denotes the number of genes and Y-axis denotes analysed fish species. Light green represents intronless gene and colour gradient increases with the number of introns.

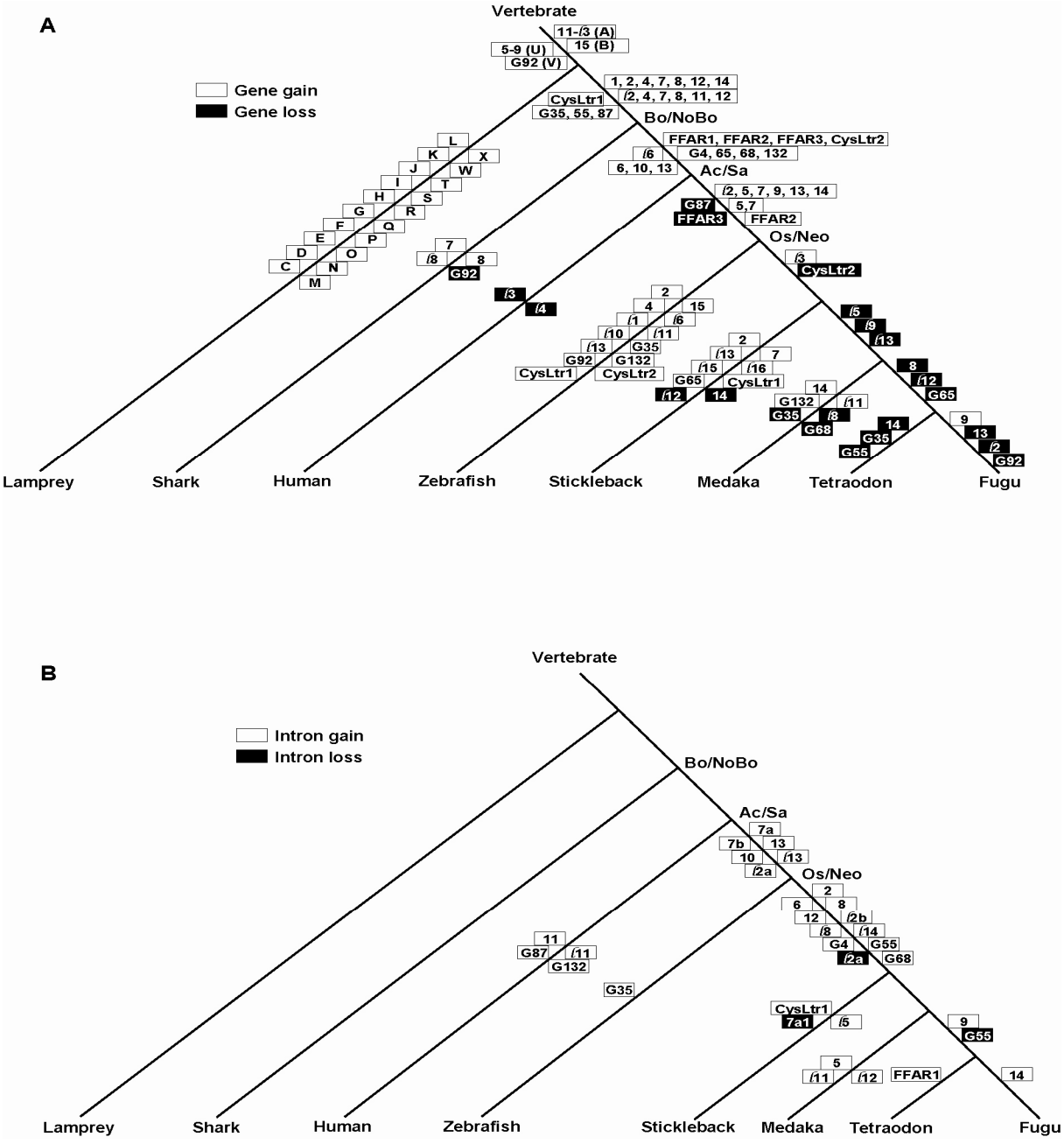


Figure 10| Estimated minimal evolutionary age and intron dynamics of P2Y subfamilies.

(A) Open squares represent gene gain events in each lineage and black squares represent gene loss events. The name of the subfamilies is indicated inside each square. (B) Open squares represent intron gain events whereas black squares represent intron loss events. The major phylogenetic transitions are indicated: Bo/NoBo, bony fish/cartilaginous fish; Ac/Sa, actinopterygian/sarcopterygian split; Os/Neo, ostariophysii/neoteleostei segregation. The number indicates P2Y subtypes; l, p2y-like; G, G-protein-coupled receptor; CysLtr, cysteine leukotriene receptor.

2.10 P2Y genes are broadly expressed in adult tissue

Nine P2Y candidates are randomly selected to be characterised. These genes are $P2Y_1$, $P2Y_4$, $p2y_5$, p2y10, FFAR2c, Gpr35b, Gpr35c, p2yl-6 and p2yl-10a. To analyse the expression pattern of the selected candidates, RT-PCR is performed on twelve adult tissues (barbel and lips, olfactory epithelium, eye, brain, gills, heart, liver, stomach, kidney, testis, spleen and skin). All the analysed candidates' transcript is detected in gills. All but one, FFAR2c, of the candidates is expressed in olfactory epithelium, brain, spleen and testis. Only Gpr35a and p2yl-10a exhibited transcripts in the skin tissue (Fig. 11). FFAR2c only have transcript in gills but not in other analysed tissues. The specific expression of FFAR2c suggests that FFAR2c may play a role in gills. The remaining of the candidates displayed a broader expression, normally in several tissues, suggesting they may involve in general role in maintenance and development. Although Gpr35a and Gpr35c are clustered together in the phylogenetic analysis, the two genes displayed a different expression pattern. Gpr35a transcript is found in all the analysed tissue. In contrast, Gpr35c transcript is expressed only in olfactory epithelium, brain, gills, stomach, kidney, testis and spleen. The differences in the expression pattern suggest that the genes might undergo partitioning of functions.

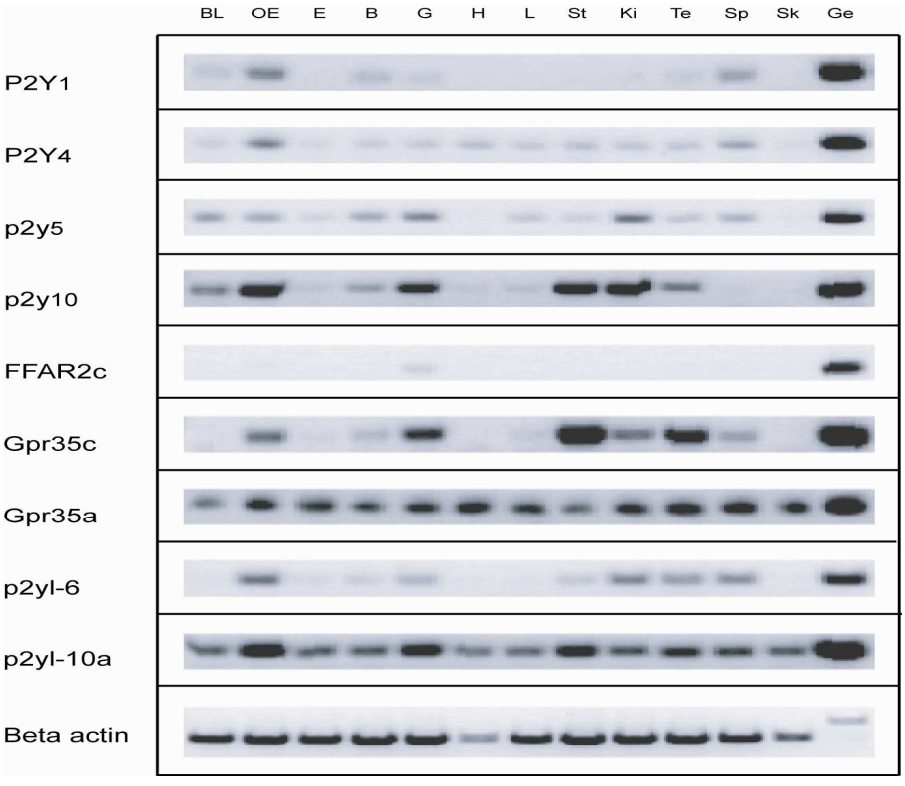


Figure 11 | Broad expression of P2Y transcripts in different types of adult tissues.

RT-PCR indicating transcription of P2Y genes are found in several adult tissues. PCR amplifications were performed by using gene specific primers that amplify between 300 to 400bps. Beta-actin was used as a quality control for the cDNA generated. Genomic-contaminated cDNA would generate 800bps band whereas genomic-free cDNA generates 500bps band. Genomic DNA (Ge) in the last lane is a positive control for the PCR. Adult tissues used are as follows: BL, barbels and lips; OE, olfactory epithelium; E, eye; B, brain; G, gills; H, heart; L, liver; St, stomach; K, kidney; T, testis; Sp, spleen; Sk, skin.

2.11 Expression of P2Y in larvae

Whole mount *in situ* hybridisation is carried out on 5 days post fertilization (dpf) zebrafish larvae. At 5dpf, the larvae have completed organogenesis and major behavioural patterns are developed. All the analysed genes were expressed in the brain and pharyngeal arches which aftermath, give rise to gills. This expression data corresponds to PCR results. Expression pattern of P2Y4 and p2yl-10a are not included as the sense probe showed similar labeling as the antisense probe albeit at lesser intensity. Sectioning is performed on larvae to analyse the expression pattern in details in particularly in regions which are not accessible by light microscope of whole-mount larval.

$P2Y_1$

$P2Y_1$ transcripts are detected in the central nervous system, otic vesicle and pharyngeal arches. Sectioning of the 5dpf larval revealed the expression in the brain and spinal cord that formed the component of central nervous system. In the brain, the expression is restricted to the ventricular regions in telencephalon, diencephalon and rhombencephalon, and midbrain hindbrain boundary (Fig. 12A-B, D-F, H-M). Similarly, $P2Y_1$ transcripts are also detected in the ventricular zone of spinal cord (Fig. 12G, K-M) that is extended from rhombencephalon, but not in floor plate. Besides that, $P2Y_1$ transcripts are also detected in pharyngeal arches. The expression is observed only at the tips of the branchial arches (Fig. 12N). A weaker expression is observed in Merkel's cartilage (Fig. 12C-D). In addition, $P2Y_1$ transcripts are also detected in pectoral fin (Fig. 12N) and otic vesicle (Fig. 12I-J).

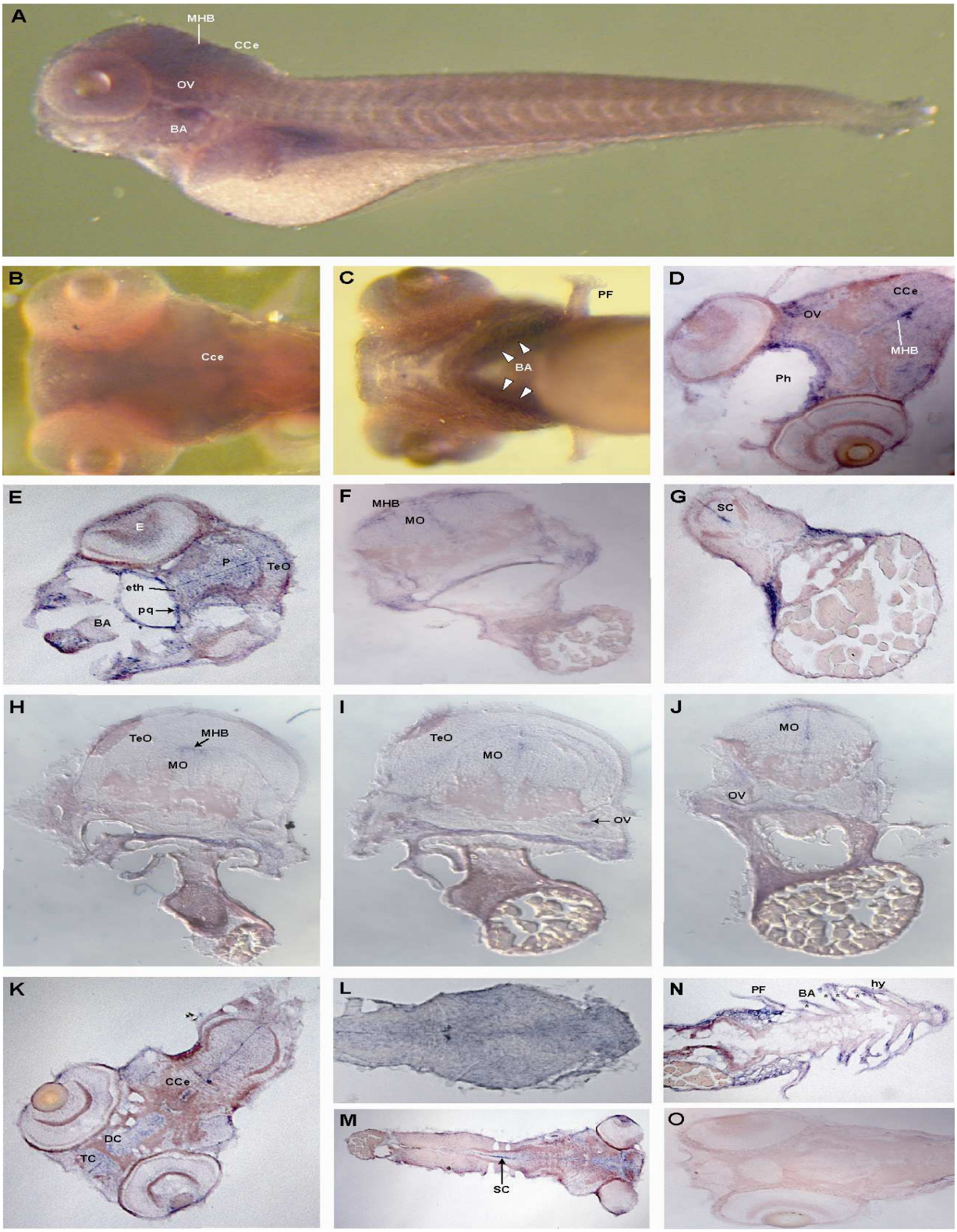


Figure 12 | Expression of $P2Y_1$ in central nervous system, otic vesicle and pharyngeal branches.

Analysis of $P2Y_1$ expression in 5dpf larval by whole mount in situ hybridization. Panel (A-C), whole mount, (D, K-O) coronal and (E-J) transverse sections after hybridization. A) Lateral view shows expression in brain including midbrain hindbrain boundary, otic vesicle and branchial arches; (B) Dorsal view shows expression in the cerebellum; (C) Ventral view shows expression in pharyngeal branches and pectoral fin; Expression is detected in the telencephalon (E, K-M), diencephalon (E, K-L) and rhombencephalon (F-M) in the brain; The expression is restricted to the ventricular region of the spinal cord (G, L-M); The edges of branchial (*) and hyoid arches are labeled as well as pectoral fin (N); (O) Sense riboprobe. *Abbreviations*: BA: branchial arches; CCe; cerebellum; CMZ, ciliary marginal zone; DC: diencephalon; E: eye; eth: ethmoid plate; FP, floor plate; Ha, habenula; Hy: ceratohyal; Ki, kidney; ; MA: mandibular arches; MC, Merkel's cartilage; MHB: midbrain hindbrain boundary; MO: medulla oblongata; OG, octaval ganglion; OV: otic vesicle; P: pallium; pq: palatoquadrate; PF: pectoral fin; SB, swimming bladder; SC: spinal cord; TC: telencephalon; TeO: optic tectum, TG, trigeminal ganglion; Th, thalamus;

$p2y_5$

$p2y_5$ is broadly expressed in brain. It shared a similar expression of $P2Y_1$ in ventricular regions in telencephalon, diencephalon, rhombencephalon and spinal cord (Fig. 13B, D, L). In additions, $p2y_5$ transcripts are also strongly detected in habenula (Fig. 13I) and midbrain hindbrain boundary (Fig. 13A, B, E, J). A weaker expression is also detected in pallium (Fig. 13F and H) and craniofacial mesenchyme (Fig. 13H) adjacent to developing cartilage tissue. Expression in peripheral nervous system is detected in trigeminal ganglion (Fig. 13J) and octaval ganglion (Fig. 13K). The former is involved in mechanosensory system and the latter in auditory system. In pharyngeal arches, $p2y_5$ expression is detected in branchial arches (Fig. 13G), palatoquadrate (Fig. 13H) and Merkel's cartilage (Fig. 13C).

$p2y_{10}$

The expression of $p2y_{10}$ is limited to dorsal pallium (Fig. 14E), midbrain hindbrain boundary (Fig. 14A), medulla oblongata (Fig. 14F) and vagal ganglion (Fig. 14A and F) in the brain. Pharyngeal region and kidney (Fig. 14G) is strongly label by the transcripts. Additionally, $p2y_{10}$ transcripts are also detected in craniofacial mesenchyme adjacent to ethmoid plate and in ciliary marginal zone of the eye (Fig. 14E).

p2yl-6

In the brain, p2yl-6 transcripts are detected in telencephalon (Fig. 15F), thalamus (Fig. 15L), proliferating ventricular regions in mesencephalon and rhombencephalon (Fig. 15F), midbrain hindbrain boundary (Fig. 15B, F) and medulla oblongata (Fig. 15M-O). Strong expression of

p2yl-6 transcripts are also observed in pallium (Fig. 15F) and vagal ganglion (Fig. 15A, N, O). Besides brain region, p2yl-6 is also expressed strongly in pharyngeal regions of palatoquadrate and ethmoid plate, craniofacial mesenchyme region (Fig. 15K), and tips of pectoral fin (Fig. 15E and O). A weaker expression is detected in tips of branchial arches, Merkel's cartilage (Fig. 15D) and otic vesicle (Fig. 15J).

Gpr35a

Gpr35a is expressed in the brain and pharyngeal regions. In the brain, the expression is distributed from anterior region of the subpallium to medulla oblongata in rhombencephalon (Fig. 16D-G, I). The expression in pharyngeal regions is broadly distributed in branchial (Fig. 16C and J) and mandibular arches with strongest expression observed in palatoquadrate (Fig. 16D). In addition, Gpr35a transcripts are labeled in kidney (Fig. 16H), vagal ganglion (Fig. 16G) and ciliary marginal zone in the eye (Fig. 16D).

Gpr35c

Gpr35c shared a similar expression as Gpr35b in 5dpf larval. In the brain, the expression of Gpr35c is detected from anterior region of the pallium (Fig. 17H) through medulla oblongata in rhombencephalon (Fig. 17F). Likewise in pharyngeal region, the transcripts are detected in brachial arches with the strongest expression observed in mandibular and hyoid arches (Fig. 17C, E). A cluster of cells are labeled in the region of craniofacial mesenchyme adjacent to trabeculae (Fig. 17D, H). Similarly to Gpr35a, the kidney is also strongly labeled (Fig. 17G). Besides that, expression is also detected in vagal ganglion (Fig. 17A and F).

FFAR2c

FFAR2c are broadly expressed in the brain (Appx. 7.2D). The expression is distributing from telencephalon to rhombencephalon (A, D-G). FFAR2c is strongly labeled in Merkel's cartilage and to a lesser extent in branchial arches (C). Expression is also detected in otic vesicle (F).

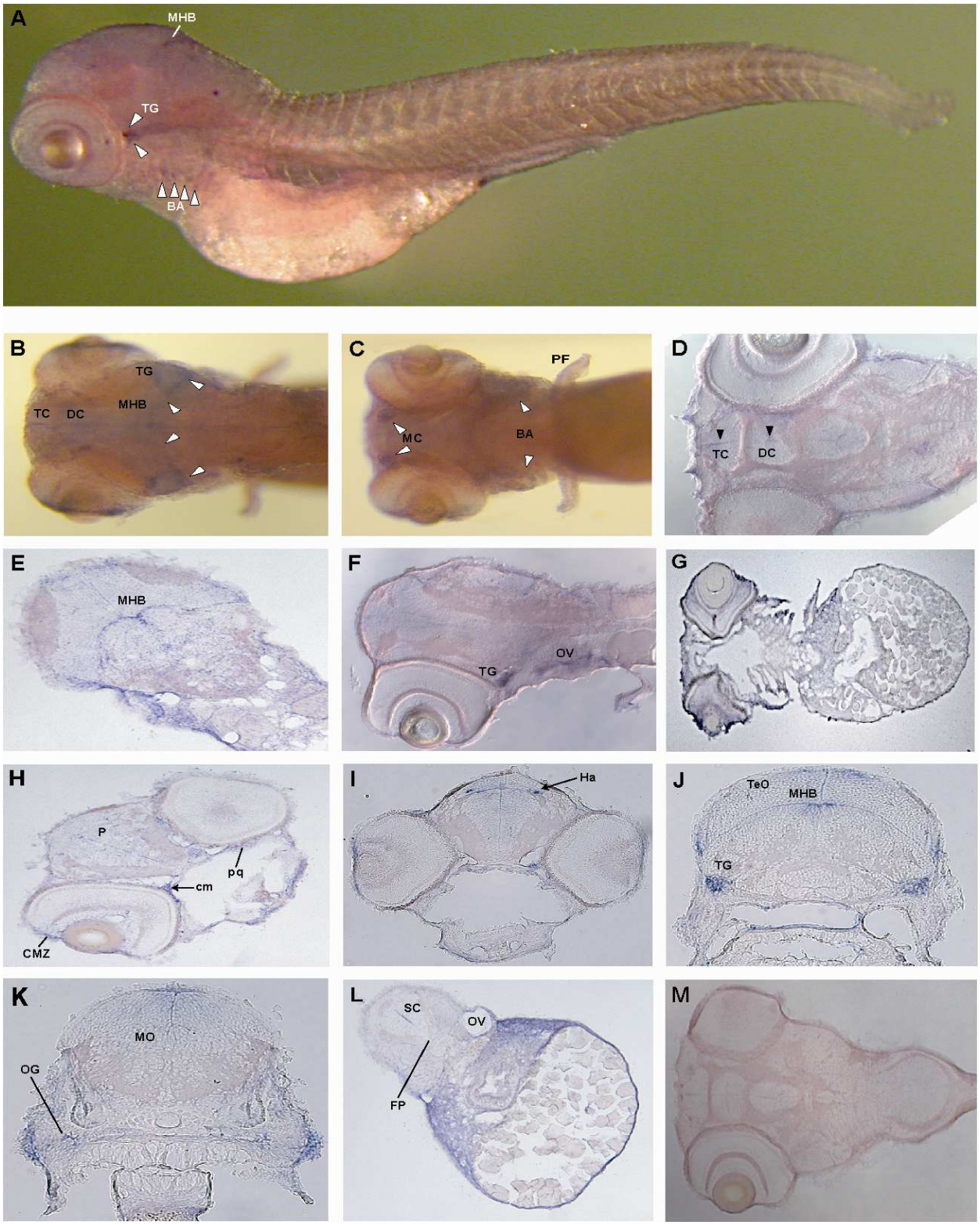


Figure 13 | Expression of $p2y_5$ in central and peripheral nervous system, auditory and mechanosensory system.

Analysis of $p2y_5$ expression in 5dpf larval by whole mount in situ hybridization. Panel (A-C), whole mount, (D-G, M) coronal and (H-L) transverse sections after hybridization. A) Lateral view shows expression in brain including midbrain hindbrain boundary, trigeminal ganglion and branchial arches; (B) Dorsal view shows expression in ventricular region of telencephalon, diencephalon and rhombencephalon, trigeminal ganglion and midbrain hindbrain boundary; (C) Ventral view shows expression in Meckel's cartilage, branchial arches and pectoral fin; In the central nervous system, expression is detected in the ventricular region of telencephalon (D, F, H), diencephalon (D, F) and rhombencephalon (E, K) in the brain and spinal cord (L); Trigeminal ganglion which is a component of menchanosensory system is labeled (J); In auditory system, expression is detected in octaval ganglion and otic vesicle (F, K). Expression is also detected in branchial arches (G), pectoral fin (F and G), craniofacial mesenchyme and ciliary marginal zone of the eye (H-I); (O) Sense riboprobe.

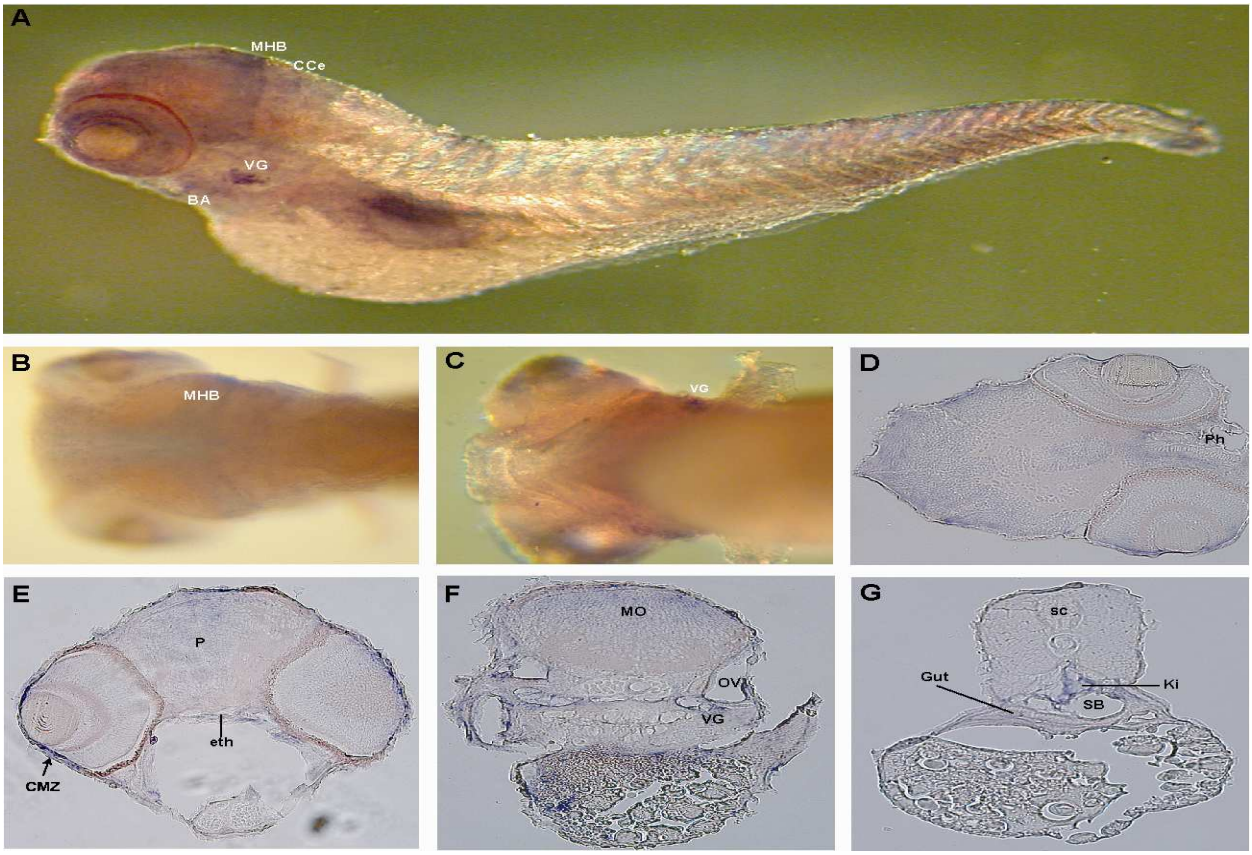


Figure 14 | Expression of $p2y_{10}$ in brain and ciliary marginal zone of the eye.

Analysis of $p2y_{10}$ expression in 5dpf larval by whole mount in situ hybridization. Panel (A-C), whole mount, (D) coronal and (E-G) transverse sections after hybridization. A) Lateral view shows expression in brain including midbrain hindbrain boundary, vagal ganglion and branchial arches; (B) Dorsal view shows expression in midbrain hindbrain boundary; (C) Ventral view shows expression in vagal ganglion and branchial arches; In the brain, the expression is detected in the dorsal of pallium, midbrain hindbrain boundary (E), dorsal of medulla oblongata and vagal ganglion (F); Expression is also detected in pharynx (D), craniofacial mesenchyme, ciliary marginal zone of the eye (E) and kidney (G).

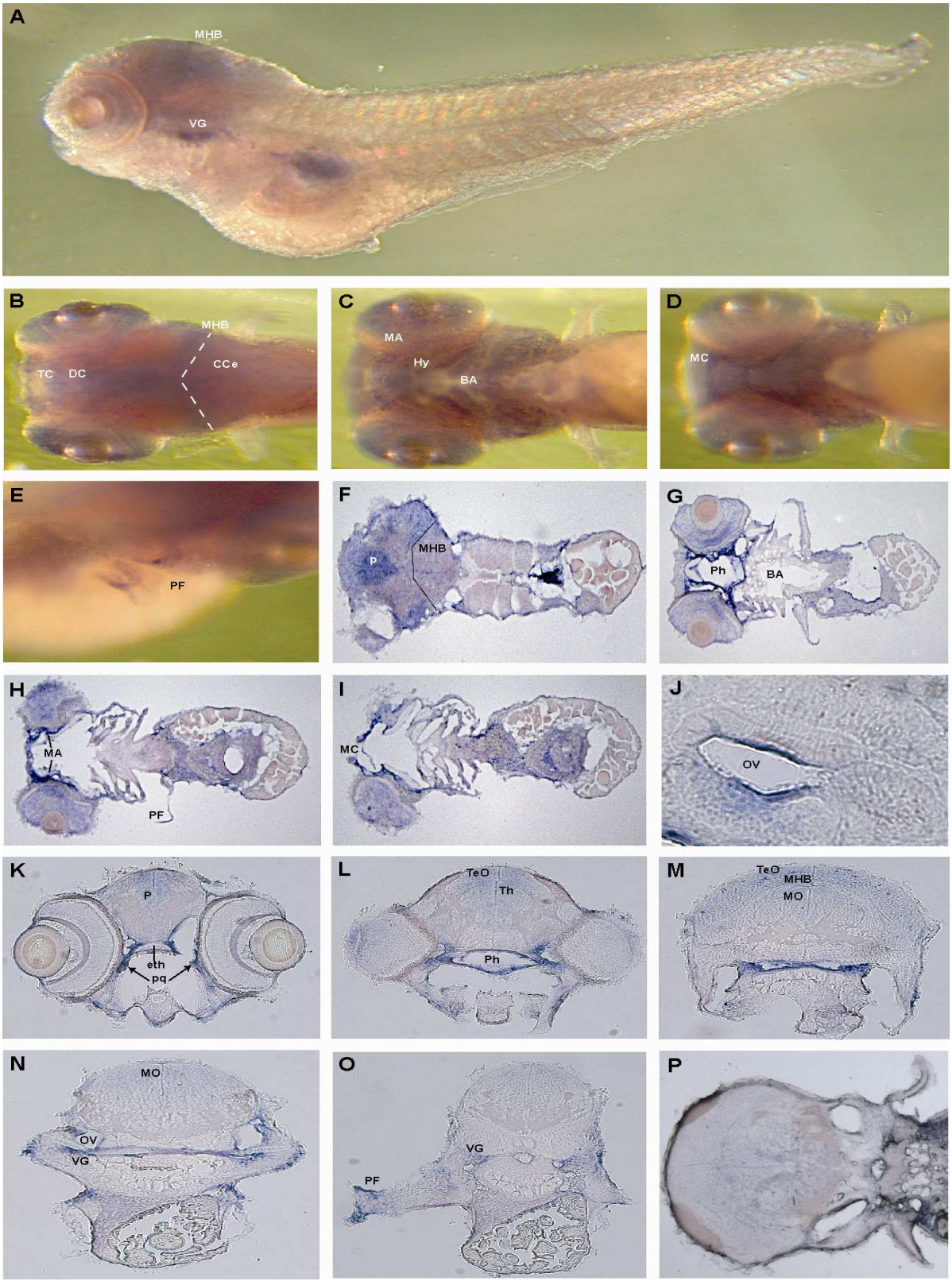


Figure 15 | Expression of p2yl-6 in brain, pharyngeal branches, otic vesicle and vagal ganglion.

Analysis of p2yl-6 expression in 5dpf larval by whole mount in situ hybridization. Panel (A-E), whole mount, (F-J) coronal and (K-O) transverse sections after hybridization. A) Lateral view shows expression midbrain hindbrain boundary and vagal ganglion; (B) Dorsal view shows expression in ventricular region of telencephalon, diencephalon, mesencephalon and rhombencephalon, and midbrain hindbrain boundary; (C-D) Ventral view shows expression in mandibular arches, ceratohyal, branchial arches and Meckel's cartilage; (E) Dorsal lateral view shows pectoral fin is labeled; In the brain, expression is detected in the ventricular region of telencephalon (B, F, K), diencephalon (B, F, L), mesencephalon (L-M) and rhombencephalon (F, M); Expression is detected in pharyngeal branches that includes mandibular aches (C, H), palatoquadrate (K), ceratohyal (C) and branchial arhes (G-I); Several cells in optic vesicle (J) and craniofacial mesenchyme (K) are also labeled; Expression is also detected in vagal ganglion (A, N, O); (P) Sense riboprobe.

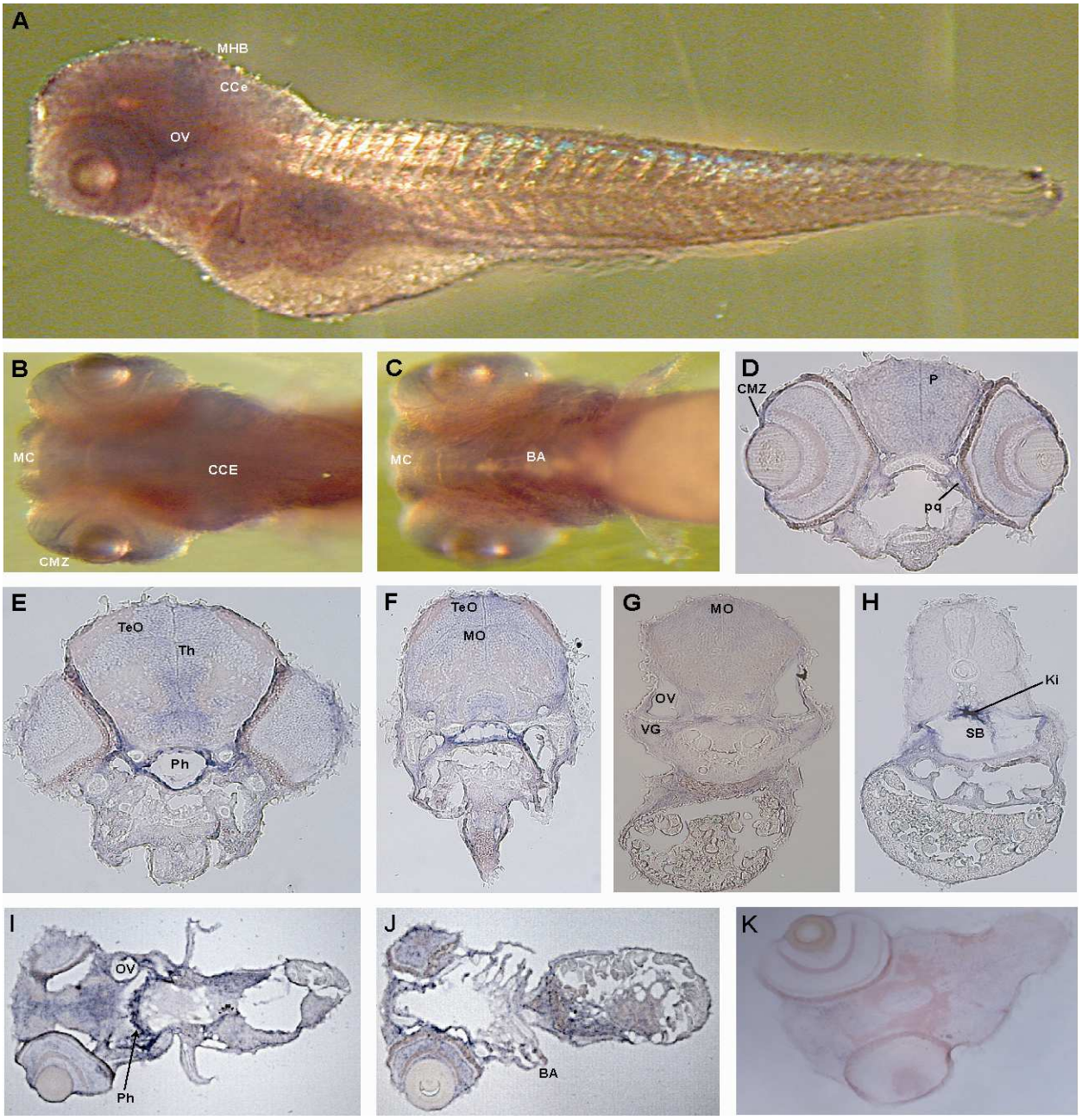


Figure 16 | Expression of Gpr35a in brain, vagal ganglion, pharyngeal arches, kidney and ciliary marginal zone in the eye.

Analysis of Gpr35a expression in 5dpf larval by whole mount in situ hybridization. Panel (A-C), whole mount, (D-H) transverse and (I-K) coronal sections after hybridization. A) Lateral view shows expression in the brain and optic vesicle; (B) Dorsal view shows expression in Merkel's cartilage, ciliary marginal zone in the eye and brain; (C) Ventral view shows expression in Merkel's artilage, ceratohyal and branchial arches. In the brain, expression is distributed from anterior region of subpallium to dorsal of medulla oblongata (D-G, I); Expression is detected in palatoquadrate and ciliary marginal zone of the eye (D), otic vesicle (I), vagal ganglion (G) and kidney (H); (K) Sense riboprobe.

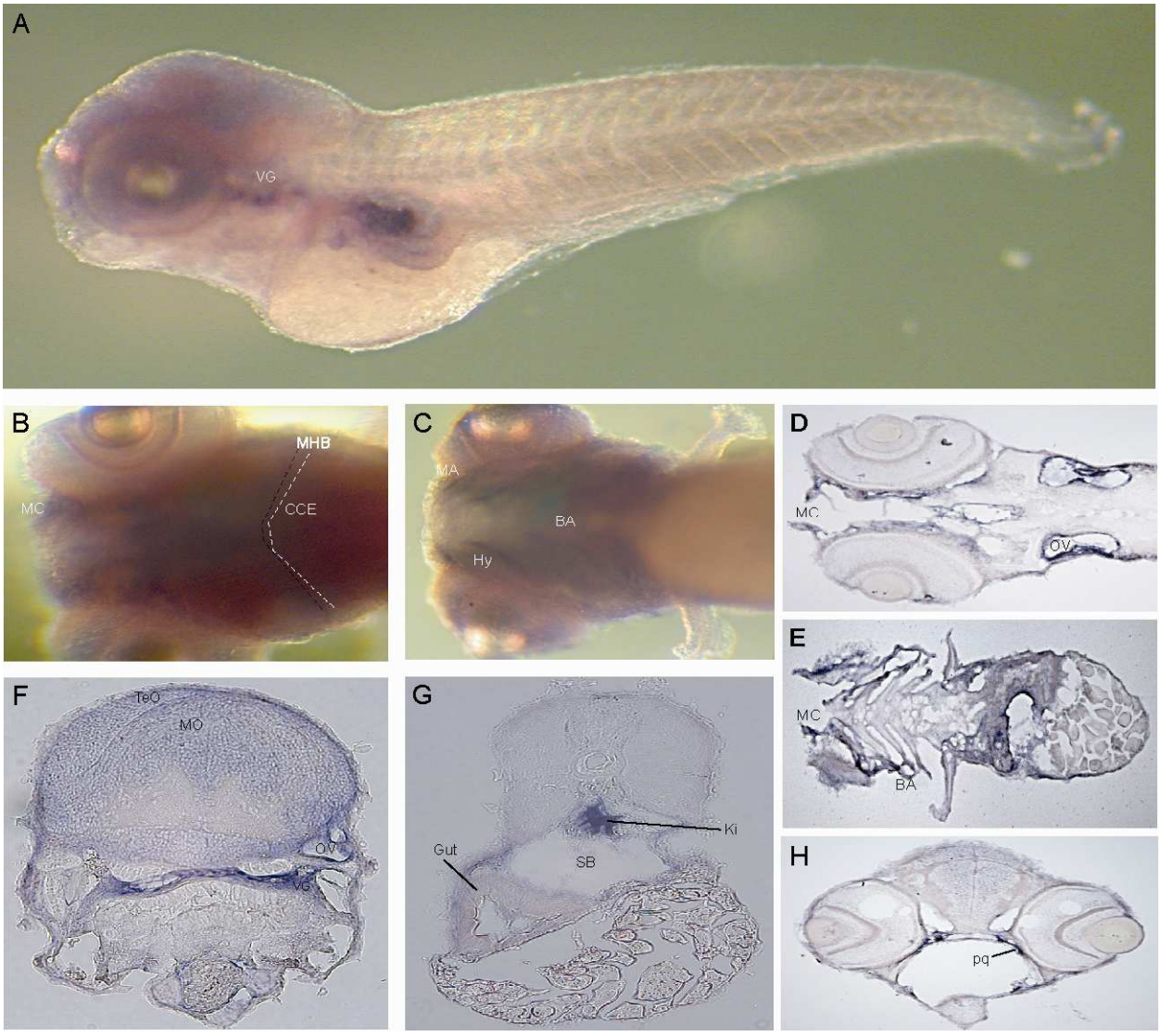


Figure 17 | Expression of Gpr35c in brain, vagal ganglion, pharyngeal arches and kidney.

Analysis of Gpr35c expression in 5dpf larval by whole mount in situ hybridization. Panel (A-C), whole mount, (D-E, H) coronal and (F-G) transverse sections after hybridization. A) Lateral view shows expression in brain and vagal ganglion; (B) Dorsal view shows expression in mesencephalon midbrain hindbrain boundary, rhombencephalon and Meckel's cartilage; (C) Ventral view shows expression in mandibular arches, ceratohyal and branchial arches; In the brain, the expression is distributed from anterior of pallium to medulla oblongata (D, H); Expression is also strongly labeled in kidney (G).

2.12 Broad expression of P2Y genes in adult olfactory epithelium

Although P2Y genes are not detected in the olfactory organ in 5dpf larval, the expression is detected in adult olfactory epithelium. The expressions are detected both in sensory and non-sensory regions (Fig. 19). This is in contrast with the expression of established olfactory receptor genes that are restricted to the sensory regions. In the sensory regions, P2Y genes are expressed throughout the lamellae, i.e. also in the regions where the known three types of olfactory sensory neurons are positioned. Hence, there is a possibility that P2Y genes are expressed in the olfactory sensory neurons or sustentacular cells. The non-sensory regions contain goblet cells, ciliated non-sensory cells with microridges which involve in regulation of water flow into olfactory epithelium. In addition, goblet cells secrete mucus upon interaction with stimuli. The secretion of mucus might influence odor detection due to the changes in mucus composition (Hansen and Zeiske, 1998; Chakrabarti and Choudhury, 2007).

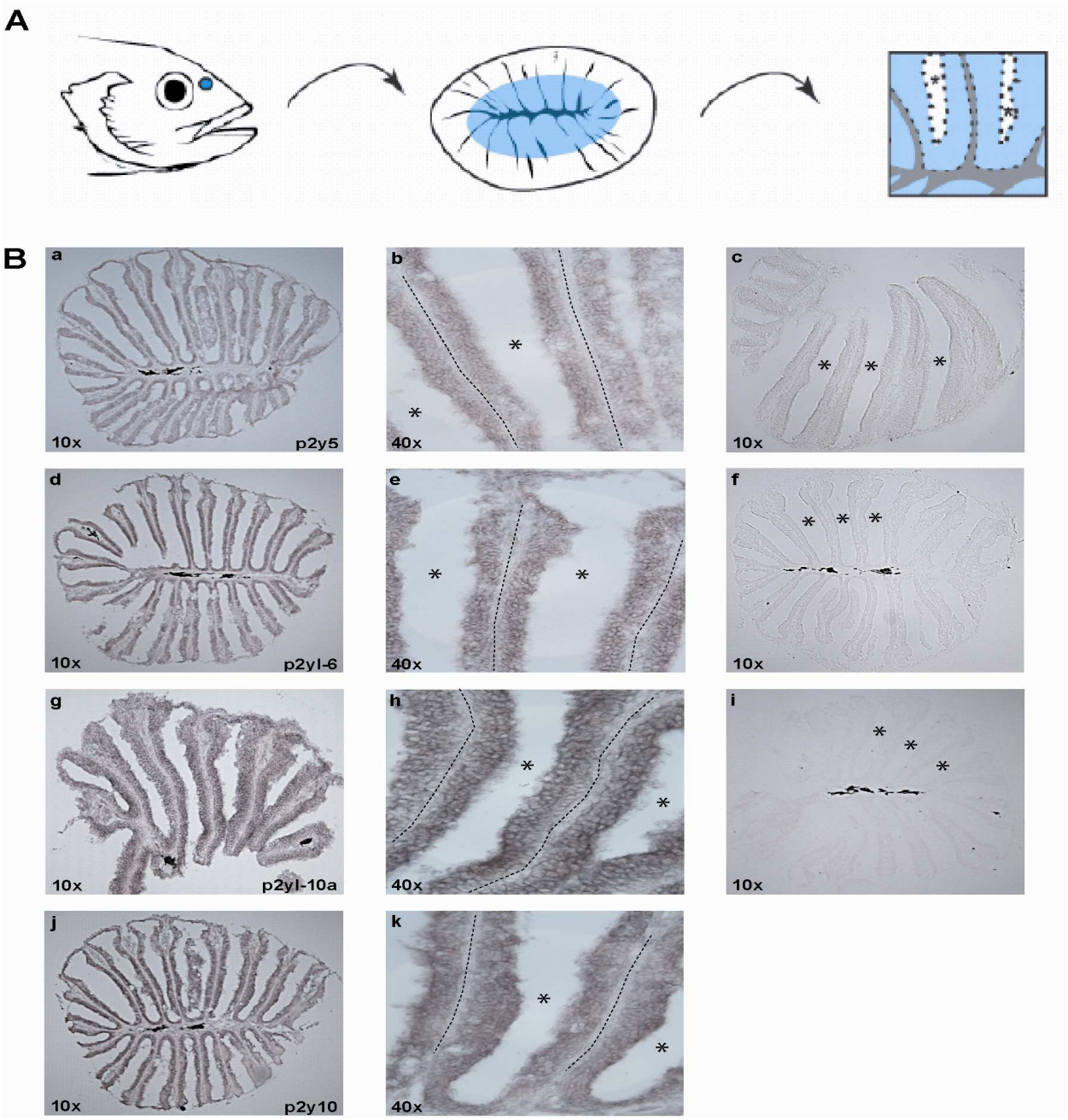


Figure 18 | Expression of P2Y genes in zebrafish olfactory epithelium.

(A) Schematic representation of the localization of the olfactory epithelium followed by a drawing of a horizontal section of olfactory epithelium (lamellae are cut perpendicular to their flat surface) and finally an enlargement of two lamellae. The central blue-coloured are in the lamellae indicates the location of sensory neuroepithelium; gray areas and thin dotted line, basal lamina; black dots and asterisk, lumen (Weth *et al*., 1996). (B) In situ hybridization with $p2y_5$, $p2y_{10}$, p2yl-6 and p2yl-10a in horizontal sections of olfactory epithelium. Panel a, b, d, e, g, h, j and k are hybridizations with antisense probes whereas panel c, f and I are with sense probes. Panel b, e, h and k show higher magnification of the lamellae. The black asterisks indicate the lumen and dotted line indicates basal lamina. Red arrowheads point to the labeled cells.

3. DISCUSSION

3.1 P2Y receptors originate early in vertebrate lineage

Stimulation of ATP increased the rate of output of contractile vacuoles in *Amoeba* (Pothier *et al*., 1984, 1987; Couillard, 1986) and *Paramecium* (Organ *et al*., 1968) suggesting the emergence of purinergic signaling early in evolution. Electrophysiology methods had been used to demonstrate the presence of ATP receptors in invertebrates (Burnstock, 1996) but no receptors has been cloned until recently. The cloning of a P2X receptor in both *Dictyostelium discoideum* and *Schistosoma mansoni* (Agboh *et al*., 2004; Fountain *et al*., 2007) for the first time identified receptors involve in mediating purinergic signaling in invertebrates. However, P2Y genes were not detected in invertebrate species (Fredriksson and Schioth, 2005). The absence of P2Y genes in invertebrates was confirmed by our database mining. Therefore, it is reasonable to suggest that P2Y receptors arise in the vertebrate lineage and in fact after the split between chordates and vertebrates.

3.2 Low resolution of basal taxa in phylogenetic analysis

In all the methods used in P2Y tree construction the early sub-branching possessed a low bootstrap value, which does not allow to pinpoint the early evolution with any degree of certainty. Moreover, these low bootstrap values suggest that these genes have diverged in an extremely short period (Fredriksson *et al*., 2003). P2Y genes emerged presumably around 550 MYA based on the existence of lamprey homologues. Functional P2Y are phylogenetically divided into two groups; group I encompasses $P2Y_1$, $P2Y_2$, $P2Y_4$, $P2Y_6$ and $P2Y_{11}$ whereas group II members includes $P2Y_{12}$, $P2Y_{13}$ and $P2Y_{14}$. In zebrafish, amino acid identity within members of the group varied between 27-46%. The orthologues generally have a higher identity in pairwise comparisons (>60%). Furthermore, the amino acid identity between two groups varied between 20-29%. The low percentage of shared identical amino acid implies that P2Y genes are very divergent.

In the phylogenetic tree analysis, P2Y genes coalesced with platelet-activating, succinate and many orphan receptors. Phylogenetic distances between platelet activating factor receptors and P2Y receptors ($P2Y_{12}$-$P2Y_{14}$) are very small. In addition, both groups of receptors are involved in haematopoiesis (von Kugelgen, 2006) suggesting that the outgoup of platelet activating factor receptors shared a common ancestor with P2Y receptors and underwent subfunctionalization during the process of evolution.

3.3 Evolutionary dynamics of the P2Y family in fish

The basal lineages (lamprey and shark) have fewer genes than the higher species reflecting the continuous expansion of the P2Y family. However, the coarse composition of the family was already established in the common ancestor of shark and bony fish, i.e. more than 530 million years (MYA) ago, since most of the teleost subclades contain an orthologous shark gene. More recent gene expansions have occurred in both the teleost and the tetrapod lineage, with a preponderance in the teleost lineage. As a result, *Sarcopterygii* lineage tends to have only half the number of genes compared with *Actinopterygii* lineage. Mostly, these gene expansions appear to have been local gene duplications. Teleost species have a higher number of genes than tetrapods due to whole genome duplication that occurred 302-404 MYA (Semon and Wolfe, 2007), at the base of the teleost fish lineage. However, this whole genome duplication does not appear to explain the increased number of teleost fish P2Y family members, because it should not lead to local duplication events. The smaller size of the P2Y family in pufferfish may be related to the genome compaction observed in this genus.

Gene duplication affects the functional relationship between duplicated gene copies. Duplicated genes can either i) lose the gene function by pseudogene formation, ii) diversify gene function by neofunctionalization or iii) partition the ancestral gene function by subfunctionalization (Lynch and Force, 2000; Lynch *et al*., 2001). My observation is that mostly the duplicated genes are conserved between the distantly related zebrafish and pufferfish, which argues against possibility i) and supports the interpretation that the duplicated genes are likely to be functional. Paralogues Gpr35a and Gpr35c shared a similar expression pattern which is consistent with the subfunctionalization model of gene duplication in teleost (Lister *et al*., 2001; McClintock *et al*., 2002; Yu *et al*., 2003). However, a decisive answer to this question will have to await functional studies of these genes.

Species-specific duplication occurred in zebrafish and generated an additional six paralogues (p2yl-13) in the major cluster in chromosome 15. These genes were created through local tandem duplication from adjacent genes. This tandem duplication is rather recent based on the close similarity to the adjacent genes in the chromosome. Other species did not undergo the equal rate of expansion in this cluster. The occurrence of gene expansion in this particular cluster might be a regulatory mechanism to enhance the fitness of the genes. In clustered

genes, genes may confer selective benefits via their ability to be co-regulated and co-amplified (Schoneberg *et al*., 2007).

P2Y orthologues are found in tetrapods, bony and cartilaginous fishes. However, not all members appear to be present in all vertebrate classes. Orthologues of p2yl-3 and p2yl-4 are not found in the human genome. These genes appear to have been lost in the mammalian lineage. Likewise, orthologues of $p2y_{15}$, p2yl-6, Gpr65 and CysLtr2 are not found in the frog genome. The selective loss of some of the receptors in some of the species implies that these receptors are less important in the corresponding species. It may be compensated by some other receptors or mechanisms to ensure that the biological system is not affected. Gpr87 is found in elephant shark, frog and human, but is absent in the teleost lineage. Since an orthologue of Gpr87 is found in elephant shark, the loss of Gpr87 must have taken place after the split of *Actinopterygii* and *Sarcopterygii* lineage, affecting only the *Actinopterygii* lineage. The gene loss might occur as a mechanism to eradicate mutated or non-functional genes from the genome.

P2Y genes in the teleost lineage generally are slow-evolving, as evidenced by a strong negative selection. Negative selection removes disadvantageous mutations that are deleterious. Interestingly, subfamilies p2yl-3, p2yl-13, $p2y_{15}$ and Gpr35 display a higher than average dN/dS ratio. Most of these subfamilies comprise duplicated genes. Gene duplication often precedes elevated evolutionary rates due to positive selection or relaxed selective constraints (Ohta, 1993). The increased in non-synonymous substitution rates in duplicate genes might be due to the fact that duplicates genes have to undergo functional specialization in order to be maintained in the genome. Hence, groups with duplicates may be under relaxed purifying selection.

A positively selected site in each of the p2yl-4 and p2yl-7 subfamilies is not located in the ligand binding domain of the receptor. This may indicate that the ligand binding domain remains unaffected, however, we cannot exclude the possibility that the positive selected sites could affect the conformational changes of the receptor to its active stage. All P2Y receptors possess four cysteine residues in the extracellular loops and N-teminal region to form two disulfide bridges. Mutation studies on cysteine residue in EC1 and EC2 in human P2Y1 established that the disulfide bridge between cysteine residues in EC1 and EC2 is critical for proper trafficking of the human P2Y1 receptor to the cell surface. Mutation studies on the second disulfide bridge

between cysteine residues of N-terminal domain and EL3 displayed impaired response of the mutated receptors. This second disulfide bridge covalently constrains the helical bundle in a circular arrangement (Hoffman *et al*., 1999). Members of p2yl-7a subfamily maintained the conserved cysteine in N-terminal domain. In contrast, in the p2yl-7b Subfamily the conserved cysteine residue is replaced by leucine, phenylalanine or serine (Fig. 5). The replacement of the conserved cysteine in all members of p2yl-7b will disrupt the formation of disulfide bridge between N-terminal domain and EL3 and therefore influence the proper functioning of the receptor. It is interesting to note that the amino acid position directly prior to the conserved cysteine in N-terminal domain in all members of p2yl-7b undergoes non-synonymous substitution. The positive selection in this site might be attributed to the compensation of the loss of conserved cysteine residue in pyl-7b members.

Twenty-three out of sixty-three genes are singletons whereas the remaining genes appear as small clusters in the zebrafish genome. The majority of the clusters comprises just two genes and only one case exists, in which 10 genes are positioned in a single cluster in chromosome 15. The close arrangement of related genes is the evidence of gene development via gene duplication (Fredriksson, 2003). The clustering of almost two thirds of the total P2Y genes suggests that most P2Y genes arise from duplication event. The clustering of P2Y genes in the genome is also perceived in the genome of stickleback, medaka and tetraodon (data not shown).

The conserved synteny of P2Y genes (p2yl-12, $P2Y_{14}$, Gpr87, $P2Y_{13}$, and $P2Y_{12}$) in human and elephant shark is also extended to the teleost lineage. Comparison of teleost fish genome (zebrafish, stickleback, medaka and tetraodon) and human genome in this locus indicates that teleost lineage possess additional genes. The genes are most likely incorporated into the locus via tandem gene duplications. Teleost genome undergoes a high level of rearrangements as reflected by the less-than-perfect synteny observed (Fig. 8a). Syntenic genes are scattered to a different numbers of chromosome depending on the fish species implying that the teleost genome is dynamic and susceptible to rearrangements.

3.4 Intron gains in the P2Y genes family

In contrast with mammalian P2Y genes which are mostly composed of monoexonic (>90%) structure, a quarter of analysed teleost P2Y genes contain up to five introns. The ancestral genomic structure appears to be monoexonic, supported by the presence of intronless

orthologues of the analysed genes in both lamprey and shark. The increased frequency of accumulated introns in teleost implies that bony fish is susceptible to intron-gain and this event is most likely to occur after the split of *Sarcopterygii* and *Actinopterygii* lineage. The accumulated introns in *Actinopterygii* lineage but not *Chondrichthyes* (ancestor of *Actinopterygii*) support the introns-late theory of intron gain in P2Y genes.

The gene structure of P2Y genes in teleost lineage is not well–conserved during evolution. The multiexonic genes are scattered in the phylogenetic tree suggesting the intron-gain event is a random process. Some subfamilies are more susceptible to accumulate introns (e.g. $P2Y_{13}$) than others (e.g. $P2Y_1$). Even within a subfamily, not all orthologues acquire introns (e.g. $P2Y_2$) and the situation became more complicated with the number of acquired introns varying between species (e.g. $p2y_{10}$).

Further analysis of the exon/intron border of P2Y genes displayed a total of 50 independent intron gain events in the teleost lineage. Twenty-two events occurred only in one fish species, i.e. late in evolution, after all five teleost species analysed here had separated from each other, five events are seen in 2 fish species, twelve events involved in 3 fish species, eight events involved in 4 fish species and only 3 events that showed all five fish species acquired introns. Zebrafish exhibited the lowest number of intron gains, only eight events. The frequency for the other four fish species, all of them *Neoteleostei*, is at least 2-fold higher than zebrafish. The results imply that neoteleost species have a higher tendency to acquire introns than zebrafish which is evolutionary more ancient than neoteleost species. It is also further indicates that some introns in P2Y genes arose recently.

It is interesting to note that all the members of p2yl-13 clade accumulated introns in the gene. It had been mentioned previously that this cluster has arisen through tandem duplications. The intron gain in this cluster might be a consequence of the tandem duplications. Venkatesh *et al*. has shown that tandem duplication of an exon or gene can give rise to novel introns through the use of cryptic splice sites (1999).

During the whole vertebrate evolution, only three events of intron loss occurred. In one of the cases, intron loss is only observed in one of the gene in the clade. All the teleost members in p2y7a clade possess two introns in their gene structure. Stickleback has two homologues in

p2y7a clade, Ga_p2y7a1 and Ga_p2y7a2. Ga_p2y7a2 maintained twon-introns structure but Ga_p2y7a1 lost one of the intron during the evolution.

Neoteleost (stickleback, medaka, tetraodon and fugu) genome size is between 385-1000Mb whereas zebrafish possess a larger genome at 1700Mb. Smaller genomes tend to have fewer introns due to the selection of genes that can produce proteins quickly in response to external stimuli (Jeffares *et al*., 2006). However, the findings for the P2Y family are not consistent with this notion, for so far unknown reasons.

3.5 Potential roles of P2Y receptors in zebrafish

Despite the fact that P2Y receptors had been identified more than 20 years ago, most of the studies were focused on pharmacological research of mammalian receptors. Only a few studies have been carried out to investigate the expression and function of P2Y receptors in embryonic development (Meyer *et al*., 1999 and Cheung *et al*., 2003). Not much is known about the functional role of these receptors in zebrafish although $P2Y_1$ has been shown to express in zebrafish thrombocytes (Gregory and Jagadeeswaran, 2002).

P2Y receptors are broadly expressed in many adult tissues and major organs as analysed by RT-PCR, suggesting that these genes play many functions in the organism. RT-PCR results showed the expression of P2Y receptors in olfactory epithelium of adult zebrafish. Conversely, no labeling is detected in olfactory organ in 5 dpf larval zebrafish suggesting that these receptors begin to be expressed later in the development. Most of the analysed P2Y receptors show redundant expression in brain, pharyngeal branches and otic vesicle in 5 dpf larval. The redundancy in expression could imply that these receptors are working in parallel with each other. The expression pattern of P2Y genes in central and peripheral nervous system, pharyngeal arches, otic vesicle, pectoral fin and kidney is consistent with the hypothesis that purinergic signaling regulates proliferation and differentiation processes in larval development.

Expression of P2Y receptors are detected in several proliferative zones in the brain. In the forebrain, expression is detected in ventricular zones in telencephalon and diencephalon, and habenula. Likewise in the mid- and hindbrain, expression is detected in midbrain hindbrain boundary (MHB), ventricular zones in hindbrain extending to spinal cord and ciliary marginal zone in the retina. The progenitor cells reside in these zones that generate neuron and glia cells. In fish, within the ciliary marginal zone (CMZ) proliferating cells are found throughout life to

support the continous growth and enlargement of the retina (Raymond *et al*., 2006). Neural progenitor cells secrete purine nucleotides. Secreted nucleotides activate P2Y receptors to mobilize intracellular calcium and suppress the differentiation into neurons and glia cells. The addition of antagonists promotes differentiation and inhibits progenitor cell expansion (Lin *et al*., 2007). Therefore, $P2Y_1$, $p2y_5$ and $p2y_{10}$ have the potential to play a role in regulating neural progenitor cell expansion and neurogenesis in zebrafish.

Nucleotides are one of the feeding cues in fish. Stimulation of taste buds with taste stimuli will induce Type II cells to release ATP, a neurotransmitter. Subsequently, nucleotide receptors, i.e. P2X and P2Y receptors are activated to mediate signaling cascade between taste buds and afferent taste fibers (Kataoka *et al*., 2004 and Yang *et al*., 2008). In fish, mandibular and branchial arches develop into jaw and gills respectively. Taste bud primordia are present on the lips, mouth, oropharyngeal cavity and gill arches after 4-5 dpf in zebrafish (Hansen *et al*., 2002). The expression of P2Y receptors are detected in branchial arches, pharynx and Merkel's cartilage in which taste bud primordia may develop. Hence, it is conceivable that these P2Y receptors may be involved in taste signaling in zebrafish.

$P2Y_1$ and $P2Y_4$ receptors are involved in mediating pain transmission in rat. The receptors are expressed in dorsal root, nodose and trigeminal ganglion (Ruan and Burnstock, 2003), responsible for sensation in the body, heart, larynx, lungs, alimentary tract and face. On the contrary, zebrafish $P2Y_1$ did not have any expression in any of these tissues in 5dpf larval. This might be due to the late onset as the expression was observed in adult rats. However, transcripts of another P2Y gene, $p2y_5$ are detected very strongly in trigeminal ganglion in zebrafish larval, indicating that this receptor plays an early role in mediating facial sensation in zebrafish. Possibly, functional roles are fluid within the P2Y family, and orthologues do not necessarily fulfil the same roles in different species.

Nucleotide signaling in the inner ear is mediated by the hair cells which function as function as mechnosensory transducer that relay signals to brain (Haddon and Lewis, 1996). P2Y regulate the electrical gradient via marginal cells of stria vascularis and vestibular dark cells (Houseley, 1998). The labeling of epidermal lining of otic vesicle by P2Y receptors suggests that P2Y receptors may be involved in auditory transmission in zebrafish.

4. Conclusion and outlook

A thorough phylogenetic analysis of the P2Y receptor family resulted in establishing 38 clades within the family. In most cases, clades correspond to orthologues of a single gene, sometimes they encompass recently duplicated paralogues. Three clades are only present in teleosts, not in tetrapods, implying gene gains early in the teleost lineage. Teleost families contain up to twofold the gene number than mammalian families, mainly as a consequence of local gene duplications in the teleost lineage, especially in zebrafish, which has the largest gene family with 68 genes. The P2Y receptor repertoire also is shaped to some extent by gene losses, two in humans and one in teleosts. Despite this evolutionary dynamic and some rearrangements in the genome, a rough synteny of genomic positions is observed for two clusters of P2Y genes from cartilaginous to teleost fish to human, presumably reflecting the functional importance of these genes. The presence of P2Y homologues in shark and even in lamprey, but not in ascidians (*Ciona*) indicates an origin early in vertebrate evolution.

A remarkable feature of the P2Y family is the accumulation of introns in many genes. Four intron gains are observed in the human lineage, but many more gains and a few intron losses occur in the teleost lineage. Some intron gain events even occured within the pufferfish genus, i.e. less than 30 million years ago. Taken together, P2Y genes show a remarkable evolutionary versatility suggesting a correspondingly dynamic evolution of function in this family.

P2Y genes are broadly expressed especially in the proliferative zone of the brain during development, supporting the importance of purinergic signaling in the development of central nervous system. Most of the genes showed redundancy in expression pattern indicating that these receptors are working in parallel with each other. All P2Y genes analysed are ubiquitously expressed in the adult olfactory epithelium, both in sensory and non-sensory regions. This expression pattern does not suggest an olfactory function for P2Y receptors, but it does not exclude it either. Further experiments need to be conducted with cell markers in order to identify the population of cells that expresses P2Y genes (olfactory sensory neurons, sustentacular cells, globular cells, nonsensory ciliated cells). If an expression in olfactory receptor neurons can be shown, nucleotide binding of heterologously expressed receptors should be attempted, followed by knockdown of endogenous P2Y receptors and subsequent imaging of the nucleotide response.

5. MATERIALS AND METHODS

5.1 Materials

5.1.1 Fish Strains

The wild-type strains used were Tü/Tü (MPI, Tübingen) and Ab/Tü

5.1.2 Bacterial Strain

Escherichia coli DH5α bacterial strain was used in transformation.

5.1.3 Plasticware

All disposable plasticware such as 15 ml and 50 ml Falcon tubes, 6-, 24-, 48-, 96-well plates, petri dishes in various sizes were from BD or Castor, purchased from Fisher Scientific or BD biosciences. 96-well plates Polyfiltronics for colony PCR were from Whatman (supplied by Fisher Scientific), 0.2 ml PCR tubes and sterile pipette tips were from M_P supplied by Fisher Scientific. Non-sterile pipette tips were supplied by LaFontaine (Forst/Bruchsal) and Labomedic (Bonn).

5.1.4 Technical Equipments

Unless stated otherwise, general lab equipment was used for the molecular and cell biology techniques, including – balances, centrifuges, electrophoresis equipment, heating blocks and plates, hybridization and incubation ovens, micropipettes, PCR and gradient thermocyclers, pH meter, shakers, sterile hood, UV transilluminator, vortexes and waterbaths. Fresh frozen sections were obtained using the Cryostat CM 1900 (Leica). A Nikon SMZ-U binocular microscope equipped with Nikon CoolPix 950 digital camera attached was used to document whole mount images. A Zeiss AxioVert microscope with an attached Diagnostic Instruments Spot-RT camera was used to document non-fluorescent images. A fluorescent microscope Zeiss Axioplan I Imaging equipped with Apotome and HRm AxioCam (Zeiss, Germany) was used to document fluorescent images of tissue in sections.

5.1.5 Chemicals

All chemicals, if not noted otherwise, were purchased from Ambion (Austin, USA), Amersham Pharmacia Biotech (Freiburg), Applichem (Darmstadt), JTBaker supplied by Fisher Scientific (Schwerte), Biozym (Hessisch Oldendorf), Calbiochem (Darmstadt), Difco (Detroit, USA), Fluka (Neu-Ulm), Merck (Darmstadt), Molecular Probes (Leiden, NL), Roth (Karlsruhe), Serva

(Heidelberg), Sigma (Deisenhofen), Roche (Mannheim, Germany) or Invitrogen (Karlsruhe, Germany) unless stated otherwise.

5.1.5.1 Chemicals and Solutions

Solutions were prepared with didistilled water. Solutions were autoclaved for 20 min at 121 bar or filter sterilized (0.2-0.45 µm pore diameter). Glassware was autoclaved and oven baked for 2 h at 180°C. For RNA-work, solutions and water were treated with 0.1% diethylpyrocarbonate (DEPC), shaked vigorously and mixed for about 20 min on a magnetic stirrer to bring the DEPC into solution. The solutions were then autoclaved to remove any trace of DEPC. Tris buffers cannot be treated with DEPC because it reacts with primary amines. DEPC decomposes rapidly into CO_2 and ethanol in the presence of Tris buffers. Therefore, Tris buffers were prepared by using water that has been treated with DEPC first. Most of the standard stock solutions like EDTA, Tris, TAE, TBE, TE, PBS, SDS, SSC, NaOAc, and culture media like LB and SOC were prepared as described in (Sambrook J 1989). All solutions used are named in the text.

E3 embryo media	5mM NaCl, 0.17mM KCl, 0.33mM $CaCl_2$, 0.33mM $MgSO_4$, 5-5% Methylene blue
Hyb^+ solution (whole mount)	50% deionized formamide, 5X SSC, 0.1% Tween 20, 0.5mg/ml torula yeast RNA, 50µg/ml heparin adjusted to pH6.0 with citric acid
Hyb^- solution (whole mount)	50% deionized formamide, 5X SSC
Blocking solution (whole mount)	2% of heat-inactivated sheep serum and 2% BSA in PBST
Hyb solution (section)	50% deionized formamide, 5X SSC, 0.4mg/ml torula yeast RNA, 0.1mg/ml baker's yeast RNA, 5X Denhardt's solution
Blocking solution (section)	0.5% blocking reagent (Roche) in MABS
MABS	150mM NaCl, 50mM maleic acid, pH 7.5
PBS	137mM NaCl, 2.7mM KCl, 4.3mM Na_2HPO_4, 1.47mM KH_2PO_4, pH7.4
PBT	0.1% Tween 20 in PBS
20X SSC	300mM NaCl, 300mM Sodium citrate, pH 7.0

5.1.6 Molecular Biological Reagents

All used restriction enzymes were purchased from New England Biolabs (NEB). DNA polymerases (taq and reverse transcriptase) and DNA ladder were purchased from Bioline. DNA ligase and RNA polymerases were purchased from Fermentas and Roche respectively. RNaseA and proteinase K were purchased either form Sigma or Roche Biochemicals (Mannheim).

5.1.6.1 Molecular Biological Kit

Name of the kit	Company	Application
QIAEX II Gel extraction kit	Qiagen	Extraction of DNA from gel
QIAquick PCR purification kit	Qiagen	Purification of PCR product
QIAGEN PCR cloning kit	Qiagen	Cloning of DNA fragment
RNeasy Mini kit	Qiagen	Purification of total RNA
DIG RNA labeling mix	Roche Diagnostics	Labeling of riboprobe with DIG
Biotin RNA labeling mix	Roche Diagnostics	Labeling of riboprobe with biotin

5.1.6.2 Cloning Vector

Cloning vector	Company
pDrive	Qiagen

5.1.6.3 Oligonucleotide Primers

Oligonucleotide primers were purchased from Operon (Cologne) or Invitrogen Life Technologies. The primers were dissolved at a standard concentration of 50mM. Working dilutions were prepared at a concentration of 5mM and stored at -20°C. Primers were used for sequencing, cloning, and preparation of probes via the addition of T3-RNA polymerase binding site. All used primers are listed as below:

- M13-Fwd, GTAAAACGACGGCCAGT,
- M13-Rev, AACAGCTATGACCATG,
- T3, TATTAACCCTCACTAAAGGGAA,
- Dr *actin*-Fwd, CCCCATTGAGCACGGTATT,

- Dr *actin*-Rev, AGCGGTTCCCATCTCCTG,
- Dr *p2y1*-Fwd, ATGACAGCGGAGTTTAATAACCTG,
- Dr *p2y1*-Rev, TCACATGCGGTTTTCACCGT,
- Dr *p2y4*-Fwd, ATGCCGATGTCTTCGAAGGA,
- Dr *p2y4*-Rev, CTAGAGTCTCGAATCACTGC,
- Dr *p2y5*-Fwd, ATGTACAATACAAGTCTTGAGA,
- Dr *p2y5*-Rev, TCATACGTGGGACTCATTGTGA
- Dr *p2y5*-Fwd, ATGACATCCATGAATACATC,
- Dr *p2y5*-Rev, TCATTCCTTATATGAAGATG,
- Dr *p2yl-6*-Fwd, TCATACGTGGGACTCATTGTGA,
- Dr *p2yl-6*-Rev, TTATTGTTCCAAACTATTCAAA,
- Dr *p2yl-11a*-Fwd, ATGAACAACTATTCTCAAAA,
- Dr *p2yl-11a*-Rev, TTATTTGATTATTATTGTTG,
- Dr *Gpr35b*-Fwd, ATGTCCAACTGCACGCTCAA,
- Dr *Gpr35b*-Rev, TTAAGACTTGCCATCAGAAT,
- Dr *Gpr35c*-Fwd, ATGAACTCCTCCAACTCATCGA,
- Dr *Gpr35c*-Rev, TTAAGTAAAACCAGTATCCACA,

5.1.6.4 Antibodies

Name	Source	Working dilution
Anti-DIG antibody	Roche Diagnostics	1:5000

5.1.6.5 Dyes, Substrates, Embedding Media and Counterstains

Types	Component		Source
Alkaline phospahatse substrates	NBT/BCIP	Blue/violet chromogenic precipitate	Roche Diagnostics
Embedding Medium	Vectamount	Chromogenic substrates	Vector

5.2 Methods

5.2.1 Embryo maintenance

Adult zebrafish (*Danio rerio*) were kept in group tanks at a day/night rhythm of 14/5 h at a water temperature of 28°C and fed daily of dry flake foods and brine shrimp (artemia; Brustmann,Oestrich-Winkel). The aquaria were filled with a one-to-one mixture of demineralized water and tap water. Zebrafish embryos and larvae were kept in petri dishes at a density of about 50 embryos/petri dish in E3 embryo medium at 28°C without feeding for five days. Afterwards, they were raised in 2l containers. They were fed with special food for fish larvae until the age of two weeks (TetraMin Mini, TETRA), followed by artermia afterwards.

In order to elicit a controlled reproductive activity, selected females and males were put into a separate tank. Early in the following morning, fertilized eggs were collected and their age was determined using the staging criteria of Kimmel et al. (1995). The embryos were then raised and collected at 24h intervals for histological and immunohistochemical processing. Embryos fixed at a stage older than 24 h postfertilization (hpf) were raised in 2 mM 1-phenyl-2-thiourea (PTU) in embryo medium after the epiboly stage (about 12 h) to prevent pigmentation. The embryonic and larval stages used for all investigations reported here ranged between one and 21 days postfertilization (dpf).

5.2.2 Data mining/Identification of P2Ys gene

Mammalian P2Y receptors were extracted from National Center for Biology Information protein database. Two strategies were employed in the subsequent data mining analysis in seven fish species: (1) the algorithm tBLASTN was applied to compare amino-acid query sequences to the DNA databases (http://www.ensembl.org'index.html) with a non-stringent cut-off value of 5^{-5} and (2) the automatically orthologue predicted genes predicted in fish species were retrieved from query sequences.

Subsequently, the P2Y-like genes are then verified through inclusion criteria. The inclusion criteria used were: (1) positioning within the P2Ys clade in phylogenetic analysis; (2) employing BLASTP algorithm in the NCBI non-redundant database in which first hits should resulted annotated P2Ys or P2Y-like candidate and had a significant E-value; (3) presence of seven transmembrane domains (based on consensus of the prediction results obtained by using TMHMM at http://www.cbs.dtu.dk/services/TMHMM and TMpred at http://www.ch.embnet.org/software/TMPRED_form.html) and (4) CDS length of between 800 to

1200bps. Duplicate genes were removed and resulting genes were subjected to the further analyses described below. The automatically annotated orthologues of fish P2Y genes in *Xenopus tropicalis* and mammalians were also retrieved and included in subsequent phylogenetic analysis.

All contigs containing P2Y receptors BLAST hits were manually collected. ORF finder (www.ncbi.nlm.nih.gov/projects/gorf/) and GeneWise (http://www.ebi.ac.uk/Wise2/index.html) programs were employed to retrieve the full-length of the sequence.

5.2.3 Phylogenetic analysis

To infer the relationship of the sequences, phylogenetic analysis of distance-based, maximum parsimony and maximum likelihood analyses were performed using Neighbour-Joining (NJ), Protpars (MP) and Proml (ML) programs respectively. Olfactory receptors are used as outgroups to root the tree. Bootstrap analysis was carried out to measure the robustness of branching patterns of the tree.

5.2.3.1 Sequence alignment

The multiple protein sequences were aligned by using E-NS-I strategy with the default parameters in MAFFT version 6 (http://align.bmr.kyushu-u.ac.jp/mafft/online/server/). The obtained alignment was then manually refined in MEGA 4, without removing the transmembrane regions. Due to the lack of sequence similarity, only short segments of the amino-terminal domain and carboxyl-terminal domain were aligned.

5.2.3.2 Neighbour-Joining method

Neighbour-Joining method was performed using ClustalX with bootstrapping 500 replicates.

5.2.3.3 Maximum parsimony method

The 387 aligments used were bootstrapped 50 repetitions using SEQBOOT from PHYLIP 3.67 package. The bootstrapped files were used for calculating maximum parsimony trees with Protpars from PHYLIP 3.67 package. The trees were unrooted and calculated using ordinary parsimony. Consensus trees were obtained using CONSENSE program of the PHYLIP package.

5.2.3.4 Maximum likelihood method

The 387 aligments used were bootstrapped 50 repetitions using SEQBOOT from PHYLIP 3.67 package. The bootstrapped files were used for calculating maximum likelihood trees with Proml from PHYLIP 3.67 package. The trees were unrooted and calculated using ordinary likelihood. Consensus trees were obtained using CONSENSE program of the phylip package.

5.2.4 Evolutionary analysis

5.2.4.1 Identity and similarity matrix

Pairwise alignments of the 68 fish P2Y-like amino acid sequences were performed using the EMBOSS Pairwise Alignment Algorithm (http://www.ebi.ac.uk/emboss/align/) and both the Identity and Similarity values from all the possible comparisons were retrieved and used to make the matrix.

5.2.4.2 Sequence logo

Sequence logos were generated using a web-based program, Weblogo (version 2.8.2) developed by (http://weblogo.berkeley.edu/logo.cgi). A logo was generated with the 227 fish P2Y amino acid sequences representing full-length ORFs. Sequence alignments were manually edited using MEGA 4 (Kumar *et al*., 2004) and highly divergent pieces between the start codon and the beginning of TM1 and 7 amino acids downstream of the conserved Proline (P) in the TM7 were trimmed to avoid N- and C-terminal length heterogeneity. This did not affect significantly conserved residues. Gap positions present in more than 85% of the sequences were deleted completely.

5.2.4.3 dN/dS analysis

The global dN/dS ratios for the full length ORF of the 227 fish P2Y receptor coding sequences were determined using Single Likelihood Ancestor Counting (SLAC) package (http://www.datamonkey.org), which implements the Suzuki-Gojobori method (Suzuki and Gojobori, 1999). The nucleotide alignment was manually edited to match the amino acid alignment used in the phylogenetic trees and sequence logo. To make inferences about selective pressure (positive and negative selection) on individual codons (sites) within the coding sequence of the teleost P2Y genes, the Single Likelihood Ancestor Counting (SLAC) package (http://www.datamonkey.org), which implements the Suzuki-Gojobori method (Suzuki and Gojobori, 1999), was used. The algorithm is briefly outlined. First, a best-fitting nucleotide substitution model was automatically selected by fitting several such substitution models to both

the data and a neighbor-joining tree generated from the alignment described above. Taking the obtained substitution rates and branch lengths as constant, a codon model was employed to fit to the data and a global dN/dS ratio was calculated. Then a codon by codon reconstruction of the ancestral sequences was performed using maximum likelihood. Afterwards the expected normalized (ES) and observed numbers (EN) of synonymous (NS) and non-synonymous (NN) substitutions were calculated for each non-constant site. dN = NN/EN and dS = NS/ES were then computed, and if dN < dS (negative selection) or dN > dS (positive selection), a p-value derived from a two-tailed extended binomial distribution was used to assess significance. Tests on simulated data (S.L.K. Pond and S.D.W. Frost, methods available at http://www.datamonkey.org) show that p values equal or smaller than 0.1 identify nearly all true positives with a false positive rate generally below the nominal p value; for actual data, the number of true positives at a given false positive rate is lower. In the present study, two thresholds for significance (0.1 and 0.2) were taken into account in order to identify residues potentially involved in odorant-binding activities.

5.2.5 Molecular biology techniques

5.2.5.1 Cloning of P2Y genes

5.2.5.1.1 Amplification of P2Y genes

Putative P2Ys were amplified from cDNA of olfactory epithelium and 5dpf larvae, and genomic DNA using the following primers: P2Y1, P2Y4, P2Y5, P2YR6, P2YR8, P2Y5, P2YR13 and P2YR17. Twenty-microlitre reaction mixture was assembled (5µl of 2X RedTaq Mix, 1µl each of forward and reverse primer, 5ng of template and sterile water). Denaturation of template was carried out at 96^0C for 2 min, followed by 35 cycles of 96^0C for 30 sec, 58^0C for 45 sec 72^0C for 1 min and a final elongation of 72^0C for 5 min. The PCR products of the expected sizes were excised from agarose gel and cloned into pDrive (Qiagen).

5.2.5.1.2 Gel electrophoresis

PCR products were loaded on 1% agarose gels containing 0.5µg/ml ethidium bromide in 1X TAE buffer and run at 50 V.

5.2.5.1.3 Purification of DNA fragments from agarose gel

The DNA fragment was purified using Qiagen's QIAEX II agarose gel extraction kit. The DNA band was excised from agarose gel. Three volumes of Buffer QX1 were added to one volume of excised gel. Ten-microlitre of QIAEX II was added into the mixture and resuspended by

vortexing. The mixture was incubated at 50^0C for 5 minutes to solubilize the agarose and facilitating the binding of the QIAEX II to DNA. The mixture was pelleted through centrifugation and washed with Buffer QX1 to remove residual agarose contaminants. Next, the pellet was washed twice with Buffer PE. Finally, the pellet was air-dried and the DNA was eluted by resuspending the pellet with 5mM Tris-Cl pH8.0.

5.2.5.1.4 Ligation into pDrive cloning vector

The ligation of PCR products into pDrive cloning vector was carried out using Qiagen's PCR cloning kit. Ligation mixture containing 2µl of PCR product, 2.5µl of ligation master mix and 0.5µl pDrive cloning vector was prepared and incubated at 16^0C for 30 min. After that, the ligation mixture was heated at 70^0C for 5 min to inactivate the enzyme activity.

5.2.5.1.5 Chemical transformation

One microlitre of ligation mixture was added into a tube of DH5α competent cells and mixed gently. The mixture was then incubated on ice for 30 min. Next, the cells were heat-shocked at 42^0C in a heating block for 1 min followed by rapid cooling on ice for 1 min. After that, 800µl of LB broth was added into the tube and incubated for 1 hour at 37^0C with shaking at 700rpm. Then, the cells were pelleted at 500g for 1 min and resuspended in 50µl of LB broth and plated onto LB plate containing ampicillin. Finally the plate was incubated at 37^0C overnight.

5.2.5.1.6 Plasmid DNA minipreparation

A single colony was picked from up from agar plate and inoculated into 3ml of LB broth containing ampicillin antibiotic. The bacteria were cultured overnight at 37^0C on a shaker at the speed of 180rpm.

About 1.5ml of the overnight culture was transferred into an eppendorf tube and the remaining was made up to 5% glycerol stock and stored at -80^0C. The cells were pelleted by centrifuging at 13,000 rpm for 1 min at room temperature. The supernatant was drained and the pellet was resuspended by vigorous pipetting or vortexing in 300µl Solution I containing Tris-EDTA-glucose buffer (25mM Tris-HCl pH 8.0, 5mM EDTA, 50mM glucose). Then, 300µl of freshly mixed Solution II (0.2M NaoH, 1% SDS) was added. The mixture was mixed well by inverting the tube. Next, 300µl of Solution III (3M with respect to potassium and 5M with respect to acetate) was added and mixed gently by inversion. The cell debris and bacterial genomic DNA was pelleted by centrifugation at 13,000 rpm for 5 min at room temperature. The supernatant was transferred

to a fresh tube 0.6 volume of isopropanol was added. The solution was mixed and then incubated at –20^{0}C for 5 min. The plasmid DNA was pelleted by centrifugation at 13,000 rpm for 5 min at 4^{0}C. The supernatant was discarded and the pellet was washed with 200µl of 70% ethanol. The pellet was allowed to air-dry and finally, the pellet was resuspended with 30µl of autoclaved distilled water and stored at 4^{0}C or –20^{0}C.

5.2.5.1.7 Restriction enzyme digestion of plasmid DNA

Digestions of plasmid DNA were performed using about 200ng of plasmid in 1x restriction enzyme digestion buffer and 6U of restriction enzyme in a total volume of 20µl. Digestion mixtures were incubated for 1-2 h at the appropriate temperatures as suggested by the manufacturer.

5.2.5.1.8 Preparation of glycerol stock

Glycerol stock was prepared by adding 200µl of sterile 50% glycerol to 800µl bacterial culture. The mixture was vortexed to ensure even dispersion of the glycerol and subsequently froze in liquid nitrogen. Afterwards, the tube was transferred to -80^{0}C for long-term storage.

5.2.5.1.9 DNA sequencing

DNA sequencing was carried at the core facility of the Institute of Genetics by Rita Lange on an ABI Prism 3730 DNA Analyzer (Applied Biosystems, USA).

For each cycle sequencing reaction, the following reagents were added into each tube: 2µl of BigDye terminator premix (ABI Prism), 3.2pmol primer, 200-500ng of purified plasmid DNA and autoclaved distilled water to a final volume of 5µl. Then, the mixture was mixed and briefly spun down. The sequencing profile used was as follows: 40 cycles at 95^{0}C for 20 sec, 50^{0}C for 15 sec and 60^{0}C for 4 min. The samples were ethanol precipitated and dried thoroughly. The dried samples were stored at -20°C in the dark until they were electrophoresed.

Sequence analysis was carried out using BLAST (Basic Local Alignment Search Tool), accessed through the Internet (http://www.ncbi.nlm.nih.gov/). Alignments of the sequences with several closely related genes were carried out using the CLUSTALW programme from Bioedit version 7.0.0.

5.2.5.2 Expression analysis of P2Ygenes

5.2.5.2.1 Isolation of total RNA

Total RNA was extracted from various tissues of the fish: barbel and lips, olfactory epithelium, olfactory bulb, eye, brain, grills, heart, liver, intestine, kidney, gonad, spleen and skin. The tissues were dissected in ice-cold Phosphate Buffered Saline (PBS) and stored in Buffer RLT (RNeasy Mini Kit, Qiagen). The tissue was homogenized using s syringe and needle. The lysate was passed through a 20-gauge (0.4mm) needle attached to a sterile plastic syringe for 5-5 times until a homogenous lysate is achieved. Then, RNA was isolated using RNeasy Mini Kit of Qiagen according to manufacturer's instructions. Next, purified RNA was treated with 4.7U of RNase-free DNase (Promega) for 30 min at 37^0C to remove contamination of genomic DNA. This was followed by phenol/chloroform extraction to remove the added DNAse. RNA was precipitated using 0.8M lithium chloride and 2.5 volumes of ice-cold absolute ethanol. Precipitation was allowed at -20^0C for 30 min or at $-8^0$0C for 5 min. Finally, RNA was pelleted and resuspended in the appropriate buffer.

5.2.5.2.2 Synthesis of first-strand DNA

Total RNA was mixed with 1µl of oligo $(dT)_{12-18}$ and made up to 11.8µl with DEPC-distilled water. The mixture was heated at 70^0C for 5 min to denature the secondary structure, followed by brief rapid cooling on ice. The contents of the tube were spun down by a brief centrifugation. Next, 4µl of 5X first-strand buffer, 2µl of 0.1M DTT, 1µl of 5mM dNTP mix, 1µl of RNaseOUT and 0.2µl of Bioscript (200U/µl, Bioline) were added into the mixture and mixed. Subsequently, the reaction mixture was incubated at 42^0C for 50 min. The reaction was inactivated by heating at 70^0C for 15 min. Finally, the mixture was diluted with 30µl of sterile, distilled water and the cDNA can be used as a template for amplification in PCR.

5.2.5.2.3 Polymerase chain reaction (PCR)

The 20µl PCR reaction contained 0.25µl of cDNA, 5µl of 2X RedTaq Mix, 1µl each of 5µM forward and reverse primers and distilled water. The thermal cycler regime was as follows: 95^0C for 2 min, followed by 35 cycles of 95^0C for 30 sec, 58^0C for 45 sec, 72^0C for 1 min and a final elongation step of 72^0C for 5 min.

5.2.5.3 *In situ* hybridization

In situ hybridisation techniques simultaneously detect and localize specific DNA or RNA sequences providing spatial information about their subcellular locations, within small

subpopulations of cells in tissue samples. It can identify sites of gene expression and tissue distribution of mRNA.

In situ hybridisation begins with the preparation of biological material and the labelling of a nucleic acid sequence to form the probe. The labelling of a nucleic acid involves the incorporation of either radioactive or non-radioactive marker, which can be detected. Both probe and material are denatured to ensure all nucleic acids are single-stranded. Then, under controlled experimental conditions, the single-stranded probe anneals or hybridises to its complementary single-stranded nuclei acid sequence in the biological material to form a new double-stranded molecule that incorporates the labelled marker. Finally, the sites of hybridisation are detected and visualized.

5.2.5.3.1 Cryosectioning

In situ whole-mount embryos were put in TissueTek (MILES, Elkhart, Indiana, USA), oriented and frozen at -20°C. The embryo was sectioned at 5µm. Sections were mounted on coated Superfrost plus slides and dried for 3 h at 55°C. Sections were mounted with mounting medium before viewing under the microscope.

5.2.5.3.2 Probe preparation

Sense and antisense RNA probes labeled with digoxigenin (DIG)- or biotin-labeled UTP were generated by *in vitro* transcription according to the manufacturer's instructions (Roche Diagnostics). After synthesis, the probes were not hydrolyzed into smaller pieces, as this treatment leads to elevated background signals.

Template DNA was generated by PCR using insert specific primers that contained T3 polymerase promoter sequence at the terminal end. Next, the amplified template was purified using PCR purification kit (Qiagen). The in vitro transcription reaction was prepared using about 200-500ng of purified PCR template, transcription buffer and DIG- or biotin-labeling mixture to a final concentration of 1x, 4U of T3 RNA Polymerase , 20U of RNAse inhibitor (Roche Diagnostics) and top up to final volume of 20µl. The reaction was incubated at 37°C for 2 hours and terminated by addition of 2µl of EDTA (200 mM, pH 8.0). The RNA transcript was ethanol precipitated and resuspended in 50µl of DEPC-treated H_2O. The quality of the probe was analyzed through agarose gel electrophoresis. Labeling efficiency was estimated using DIG quantification teststrips (Roche Diagnostics).

5.2.5.3.3 Preparation of embryos

Eggs were collected from mating pairs and unfertilized eggs were discarded. Embryos were allowed to develop in regular fish water until the end of gastrulation (5hpf). After that, the fish water was replaced with a solution containing 0.0045% 1-phenyl-2-thiourea to prevent pigmentation. The solution was replaced everyday until the larvae reached desired age/stage. Next, the larvae were fixed in 4% paraformaldehyde (PFA) in 1X PBS overnight at 4^0C. The fixed larvae were then dehydrated in 50% methanol for 15 min at room temperature and finally stored at -20^0C until further use.

5.2.5.3.4 Whole-mount in situ hybridization of embryos

Whole mount RNA *in situ* hybridization of zebrafish embryos or larvae was performed following the method of Thisse *et al*., (1993).

On the first day, the larvae went through a serial of rehydration in the following manner: 75% methanol: 25% PBS, 50% methanol: 50% PBS and 25% methanol: 75% PBS for 5 minutes each. Next, the larvae were washed four times with PBT, 5 mins for each washing. After that, the larvae were digested with proteinase K (5µg/ml) for 30 minutes at 37^0C to increase probe permeability into the larvae. The digestion was terminated by incubation in 4% paraformaldehyde in 1x PBS for 20 minutes. This was followed by five washings in PBT, 5 minutes for each washing. Subsequently, prehybridization was performed by incubating the larvae in 500µl of hybridization mix at 70^0C waterbath for 1 hour. After that, the prehybridization mix was replaced with hybridization mix containg about 50ng of antisense DIG-labeled RNA probe. The hybridization was left overnight (at least 16 hours) in 70^0C waterbath.

On the second day, the hybridization buffer was removed and underwent successive washing solutions to facilitate solution changes from hybridization buffer to 2x SSC. The serial high stringency washings were conducted in the following manner for 15 mins each at 70^0C: 75%HM^- : 25% 2x SSC, 50%HM^-: 50% 2x SSC, 25%HM^-: 75% 2x SSC and finally in 2x SSC. In addition, two washings of 30 mins each in 0.2x SSC were performed at 70^0C to remove non-specifically hybridised probes. Next, the larvae were progressively washed from 0.2x SSC to PBT at room temperature in the following order: 75% 0.2x SSC: 25% PBT, 50% 0.2x SSC: 50% PBT, 25% 0.2x SSC: 75% PBT and finally in PBT. All the washings were performed for 5 minutes each. After that, the larvae were blocked in blocking bluffer containing 2% sheep serum and 2mg/ml BSA for 3 to 4 hours in room temperature. Then, the blocking buffer was replaced with antibody

solution diluted at 1:5000 in blocking buffer. The incubation with antibody was left overnight with slow agitation at 4^0C.

On the third day, the antibody was removed and the larvae were washed briefly in PBT. This was followed by an extensive wash at room temperature, six times, in PBT for 15 mins each wqashing. After the last wash, the larvae were incubated at room temperature in alkaline Tris buffer (50mM Tris-HCl pH9.5, 50 mM $MgCl_2$, 50mM NaCl, 0.1% Tween 20). After that, the larvae were incubated in staining solution containing NBT-BCIP for signal detection. The staining reaction is monitored regularly until signal is perceived as sufficient. The, the staining reaction is stopped by washing the larvae with stop solution (1x PBS pH5.5, 1mM EDTA, 0.1% Tween 20).

5.2.5.3.5 *In situ* hybridization on sections of olfactory epithelia

Sections (10 µm) were fixed in 4% paraformaldehyde for 10 min at room temperature. Hybridizations were performed overnight at 60°C using standard protocols as previously described (Weth et al., 1996). Anti-DIG primary antibody coupled to alkaline phosphatase (Roche Molecular Biochemicals) and NBT-BCIP (Roche Molecular Biochemicals) was used for signal detection.

6. References

Abbracchio, M. P., G. Burnstock, et al. (2006). "International Union of Pharmacology LVIII: update on the P2Y G protein-coupled nucleotide receptors: from molecular mechanisms and pathophysiology to therapy." Pharmacol Rev **58**(3): 281-341.

Agboh, K. C., T. E. Webb, et al. (2004). "Functional characterization of a P2X receptor from Schistosoma mansoni." J Biol Chem **279**(40): 41650-7.

Alioto, T. S. and J. Ngai (2005). "The odorant receptor repertoire of teleost fish." BMC Genomics **6**: 173.

Bean, B. P. (1992). "Pharmacology and electrophysiology of ATP-activated ion channels." Trends Pharmacol Sci **13**(3): 87-90.

Bjarnadottir, T. K., D. E. Gloriam, et al. (2006). "Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse." Genomics **88**(3): 263-73.

Bockaert, J. and J. P. Pin (1999). "Molecular tinkering of G protein-coupled receptors: an evolutionary success." Embo J **18**(7): 1723-9.

Brake, A. J. and D. Julius (1996). "Signaling by extracellular nucleotides." Annu Rev Cell Dev Biol **12**: 519-41.

Burnstock, G. (1996). "Purinoceptors: Ontogeny and Phylogeny." Drug Development Research **39**: 204-242.

Burnstock, G. (2006). "Historical review: ATP as a neurotransmitter." Trends Pharmacol Sci **27**(3): 166-76.

Burnstock, G. (2006). "Purinergic signalling--an overview." Novartis Found Symp **276**: 26-48; discussion 48-57, 275-81.

Burnstock, G. (2007). "Physiology and pathophysiology of purinergic neurotransmission." Physiol Rev **87**(2): 659-797.

Burnstock, G. (2007). "Purine and pyrimidine receptors." Cell Mol Life Sci **64**(12): 1471-83.

Burnstock, G. (2008). "Purinergic signalling and disorders of the central nervous system." Nat Rev Drug Discov **7**(7): 575-90.

Burnstock, G. (2008). "Unresolved issues and controversies in purinergic signalling." J Physiol **586**(14): 3307-12.

Bush, C. F., S. V. Jones, et al. (2007). "Specificity of olfactory receptor interactions with other G protein-coupled receptors." J Biol Chem **282**(26): 19042-51.

Cheung, K. K., M. Ryten, et al. (2003). "Abundant and dynamic expression of G protein-coupled P2Y receptors in mammalian development." Dev Dyn **228**(2): 254-66.

Ciana, P., M. Fumagalli, et al. (2006). "The orphan receptor GPR17 identified as a new dual uracil nucleotides/cysteinyl-leukotrienes receptor." Embo J **25**(19): 4615-27.

Costanzi, S., L. Mamedova, et al. (2004). "Architecture of P2Y nucleotide receptors: structural comparison based on sequence analysis, mutagenesis, and homology modeling." J Med Chem **47**(22): 5393-404.

Couillard, P. (1986). "[Osmotic homeostasis in the freshwater ameba]." J Physiol (Paris) **81**(2): 132-9.

Czesnik, D., J. Kuduz, et al. (2006). "ATP activates both receptor and sustentacular supporting cells in the olfactory epithelium of Xenopus laevis tadpoles." Eur J Neurosci **23**(1): 119-28.

Derby, C. D. and P. W. Sorensen (2008). "Neural processing, perception, and behavioral responses to natural chemical stimuli by fish and crustaceans." J Chem Ecol **34**(7): 898-914.

Diogo, R., Y. Hinits, et al. (2008). "Development of mandibular, hyoid and hypobranchial muscles in the zebrafish: homologies and evolution of these muscles within bony fishes and tetrapods." BMC Dev Biol **8**: 24.

Dubyak, G. R. and C. el-Moatassim (1993). "Signal transduction via P2-purinergic receptors for extracellular ATP and other nucleotides." Am J Physiol **265**(3 Pt 1): C577-606.

Durand, D. and R. Hoberman (2006). "Diagnosing duplications--can it be done?" Trends Genet **22**(3): 156-64.

Erlinge, D. and G. Burnstock (2008). "P2 receptors in cardiovascular regulation and disease." Purinergic Signal **4**(1): 1-20.

Evans, D. H., P. M. Piermarini, et al. (2005). "The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste." Physiol Rev **85**(1): 97-177.
Fedorov, A., S. Roy, et al. (2003). "Mystery of intron gain." Genome Res **13**(10): 2236-41.
Fields, R. D. and G. Burnstock (2006). "Purinergic signalling in neuron-glia interactions." Nat Rev Neurosci **7**(6): 423-36.
Fischer, W. and U. Krugel (2007). "P2Y receptors: focus on structural, pharmacological and functional aspects in the brain." Curr Med Chem **14**(23): 2429-55.
Foord, S. M., T. I. Bonner, et al. (2005). "International Union of Pharmacology. XLVI. G protein-coupled receptor list." Pharmacol Rev **57**(2): 279-88.
Fountain, S. J., K. Parkinson, et al. (2007). "An intracellular P2X receptor required for osmoregulation in Dictyostelium discoideum." Nature **448**(7150): 200-3.
Fredriksson, R., M. C. Lagerstrom, et al. (2003). "The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints." Mol Pharmacol **63**(6): 1256-72.
Fredriksson, R. and H. B. Schioth (2005). "The repertoire of G-protein-coupled receptors in fully sequenced genomes." Mol Pharmacol **67**(5): 1414-25.
Fridmanis, D., R. Fredriksson, et al. (2007). "Formation of new genes explains lower intron density in mammalian Rhodopsin G protein-coupled receptors." Mol Phylogenet Evol **43**(3): 864-80.
Fujita, T., H. Tozaki-Saitoh, et al. (2008). "P2Y(1) receptor signaling enhances neuroprotection by astrocytes against oxidative stress via IL-6 release in hippocampal cultures." Glia.
Gayle, S. and G. Burnstock (2005). "Immunolocalisation of P2X and P2Y nucleotide receptors in the rat nasal mucosa." Cell Tissue Res **319**(1): 27-36.
Gerevich, Z. and P. Illes (2004). "P2Y receptors and pain transmission." Purinergic Signal **1**(1): 3-10.
Gloriam, D. E., T. K. Bjarnadottir, et al. (2005). "The repertoire of trace amine G-protein-coupled receptors: large expansion in zebrafish." Mol Phylogenet Evol **35**(2): 470-82.
Haddon, C. and J. Lewis (1996). "Early ear development in the embryo of the zebrafish, Danio rerio." J Comp Neurol **365**(1): 113-28.
Hansen, A., K. Reutter, et al. (2002). "Taste bud development in the zebrafish, Danio rerio." Dev Dyn **223**(4): 483-96.
Hansen, A., S. H. Rolen, et al. (2003). "Correlation between olfactory receptor cell type and function in the channel catfish." J Neurosci **23**(28): 9328-39.
Hansen, A. and E. Zeiske (1998). "The peripheral olfactory organ of the zebrafish, Danio rerio: an ultrastructural study." Chem Senses **23**(1): 39-48.
Hegg, C. C., D. Greenwood, et al. (2003). "Activation of purinergic receptor subtypes modulates odor sensitivity." J Neurosci **23**(23): 8291-301.
Hoffmann, C., S. Moro, et al. (1999). "The role of amino acids in extracellular loops of the human P2Y1 receptor in surface expression and activation processes." J Biol Chem **274**(21): 14639-47.
Holzschuh, J., N. Wada, et al. (2005). "Requirements for endoderm and BMP signaling in sensory neurogenesis in zebrafish." Development **132**(16): 3731-42.
Ishii, K. J. and S. Akira (2008). "Potential link between the immune system and metabolism of nucleic acids." Curr Opin Immunol.
Jeffares, D. C., T. Mourier, et al. (2006). "The biology of intron gain and loss." Trends Genet **22**(1): 16-22.
Joo, J. Y., B. W. Kim, et al. (2007). "Activation of NMDA receptors increases proliferation and differentiation of hippocampal neural progenitor cells." J Cell Sci **120**(Pt 8): 1358-70.
Kataoka, S., T. Toyono, et al. (2004). "Expression of P2Y1 receptors in rat taste buds." Histochem Cell Biol **121**(5): 419-26.
Kimmel, C. B., W. W. Ballard, et al. (1995). "Stages of embryonic development of the zebrafish." Dev Dyn **203**(3): 253-310.
Koles, L., Z. Gerevich, et al. (2008). "Interaction of P2 purinergic receptors with cellular macromolecules." Naunyn Schmiedebergs Arch Pharmacol **377**(1): 1-33.
Kroeze, W. K., D. J. Sheffler, et al. (2003). "G-protein-coupled receptors at a glance." J Cell Sci **116**(Pt 24): 4867-9.
Kumar, S., M. Nei, et al. (2008). "MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences." Brief Bioinform **9**(4): 299-306.

Lagerstrom, M. C. and H. B. Schioth (2008). "Structural diversity of G protein-coupled receptors and significance for drug discovery." Nat Rev Drug Discov **7**(4): 339-57.
Lee, S. Y., R. A. Nicholas, et al. (2004). "P2Y(1) receptor modulation of endogenous ion channel function in Xenopus oocytes: Involvement of transmembrane domains." Purinergic Signal **1**(1): 75-81.
Li, J., J. A. Mack, et al. (2005). "Early development of functional spatial maps in the zebrafish olfactory bulb." J Neurosci **25**(24): 5784-95.
Lin, J. H., T. Takano, et al. (2007). "Purinergic signaling regulates neural progenitor cell expansion and neurogenesis." Dev Biol **302**(1): 356-66.
Lister, J. A., J. Close, et al. (2001). "Duplicate mitf genes in zebrafish: complementary expression and conservation of melanogenic potential." Dev Biol **237**(2): 333-44.
Lynch, M. (2002). "Intron evolution as a population-genetic process." Proc Natl Acad Sci U S A **99**(9): 6118-23.
Lynch, M. and A. Force (2000). "The probability of duplicate gene preservation by subfunctionalization." Genetics **154**(1): 459-73.
Lynch, M., M. O'Hely, et al. (2001). "The probability of preservation of a newly arisen gene duplicate." Genetics **159**(4): 1789-804.
Mahler, J. and W. Driever (2007). "Expression of the zebrafish intermediate neurofilament Nestin in the developing nervous system and in neural proliferation zones at postembryonic stages." BMC Dev Biol **7**: 89.
Majumder, P., C. A. Trujillo, et al. (2007). "New insights into purinergic receptor signaling in neuronal differentiation, neuroprotection, and brain disorders." Purinergic Signal **3**(4): 317-31.
Marteau, F., E. Le Poul, et al. (2003). "Pharmacological characterization of the human P2Y13 receptor." Mol Pharmacol **64**(1): 104-12.
McClintock, J. M., M. A. Kheirbek, et al. (2002). "Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention." Development **129**(10): 2339-54.
Meyer, A. and M. Schartl (1999). "Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions." Curr Opin Cell Biol **11**(6): 699-704.
Meyer, M. P., J. D. Clarke, et al. (1999). "Selective expression of purinoceptor cP2Y1 suggests a role for nucleotide signalling in development of the chick embryo." Dev Dyn **214**(2): 152-8.
Ohta, T. (1993). "Pattern of nucleotide substitutions in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication." Genetics **134**(4): 1271-6.
Organ, A. E., E. C. Bovee, et al. (1968). "Adenosine triphosphate acceleration of the nephridial apparatus of Paramecium multimicronucleatum." J Protozool **15**(1): 173-6.
Panopoulou, G. and A. J. Poustka (2005). "Timing and mechanism of ancient vertebrate genome duplications -- the adventure of a hypothesis." Trends Genet **21**(10): 559-67.
Piotrowski, T., D. G. Ahn, et al. (2003). "The zebrafish van gogh mutation disrupts tbx1, which is involved in the DiGeorge deletion syndrome in humans." Development **130**(20): 5043-52.
Pothier, F., Forget, J., Sullivan, R and Couillard, P (1987). "ATP and contractile vacuole in *Amoeba proteus*: Mechanism of action of exogenous ATP and related nucleotides." J Exp Zool **243**: 379-387.
Pothier, F., P. Couillard, et al. (1984). "ATP and the autonomy of the contractile vacuole in Amoeba proteus." J Exp Zool **230**(2): 211-8.
Ralevic, V. and G. Burnstock (1998). "Receptors for purines and pyrimidines." Pharmacol Rev **50**(3): 413-92.
Raymond, P. A., L. K. Barthel, et al. (2006). "Molecular characterization of retinal stem cells and their niches in adult zebrafish." BMC Dev Biol **6**: 36.
Resende, R. R., P. Majumder, et al. (2007). "P19 embryonal carcinoma cells as in vitro model for studying purinergic receptor expression and modulation of N-methyl-D-aspartate-glutamate and acetylcholine receptors during neuronal differentiation." Neuroscience **146**(3): 1169-81.
Rodriguez-Trelles, F., R. Tarrio, et al. (2006). "Origins and evolution of spliceosomal introns." Annu Rev Genet **40**: 47-76.
Rompler, H., C. Staubert, et al. (2007). "G protein-coupled time travel: evolutionary aspects of GPCR research." Mol Interv **7**(1): 17-25.
Roy, S. W. and D. Penny (2007). "On the incidence of intron loss and gain in paralogous gene families." Mol Biol Evol **24**(8): 1579-81.

Ruan, H. Z. and G. Burnstock (2003). "Localisation of P2Y1 and P2Y4 receptors in dorsal root, nodose and trigeminal ganglia of the rat." Histochem Cell Biol **120**(5): 415-26.
Schoneberg, T., T. Hermsdorf, et al. (2007). "Structural and functional evolution of the P2Y(12)-like receptor group." Purinergic Signal **3**(4): 255-68.
Schulz, A. and T. Schoneberg (2003). "The structural evolution of a P2Y-like G-protein-coupled receptor." J Biol Chem **278**(37): 35531-41.
Semon, M. and K. H. Wolfe (2007). "Rearrangement rate following the whole-genome duplication in teleosts." Mol Biol Evol **24**(3): 860-7.
Shin, C. Y., K. C. Kim, et al. (2000). "Inhaled ATP causes mucin release from goblet cells of intact rats." Exp Lung Res **26**(1): 1-11.
Sikorski, Z. K., A; Burt, JR (1990). Postharvest biochemical and microbial changes. Seafood,: Resources, nutritional composition and preservation. Z. Sikorski. Florida, CRC Press**:** 55-75.
Stone, T. W. (1982). "Cell-membrane receptors for purines. Review." Biosci Rep **2**(2): 77-90.
Suzuki, Y. and T. Gojobori (1999). "A method for detecting positive selection at single amino acid sites." Mol Biol Evol **16**(10): 1315-28.
Thisse, C., B. Thisse, et al. (1993). "Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos." Development **119**(4): 1203-15.
Venkatesh, B., E. F. Kirkness, et al. (2007). "Survey sequencing and comparative analysis of the elephant shark (Callorhinchus milii) genome." PLoS Biol **5**(4): e101.
Venkatesh, B., Y. Ning, et al. (1999). "Late changes in spliceosomal introns define clades in vertebrate evolution." Proc Natl Acad Sci U S A **96**(18): 10267-71.
Volff, J. N. (2005). "Genome evolution and biodiversity in teleost fish." Heredity **94**(3): 280-94.
von Kugelgen, I. (2006). "Pharmacological profiles of cloned mammalian P2Y-receptor subtypes." Pharmacol Ther **110**(3): 415-32.
von Kugelgen, I. and A. Wetter (2000). "Molecular pharmacology of P2Y-receptors." Naunyn Schmiedebergs Arch Pharmacol **362**(4-5): 310-23.
Wang, H. (2008). "Comparative analysis of period genes in teleost fish genomes." J Mol Evol **67**(1): 29-40.
White, P. J., T. E. Webb, et al. (2003). "Characterization of a Ca2+ response to both UTP and ATP at human P2Y11 receptors: evidence for agonist-specific signaling." Mol Pharmacol **63**(6): 1356-63.
Yang, R., Kinnamon, JC., Montoya, AM. and Thomas, SM. (2008, July). Immunocytochemical and ultrastrutural studies of purinergic signaling in rodent taste buds. Poster session presented at the ISOT Meeting 2008, San Francisco, USA.
Yu, W. P., S. Brenner, et al. (2003). "Duplication, degeneration and subfunctionalization of the nested synapsin-Timp genes in Fugu." Trends Genet **19**(4): 180-3.
Yu, W. P., V. Rajasegaran, et al. (2008). "Elephant shark sequence reveals unique insights into the evolutionary history of vertebrate genes: A comparative analysis of the protocadherin cluster." Proc Natl Acad Sci U S A **105**(10): 3819-24.
Zimmer-Faust, R. G., RA; Carr, WES (1988). "The behavioral response of spiny lobsters to ATP: Evidence for mediation by P2 -like chemosensory receptors." Biol Bull **175**: 167-174.

7. APPENDIX

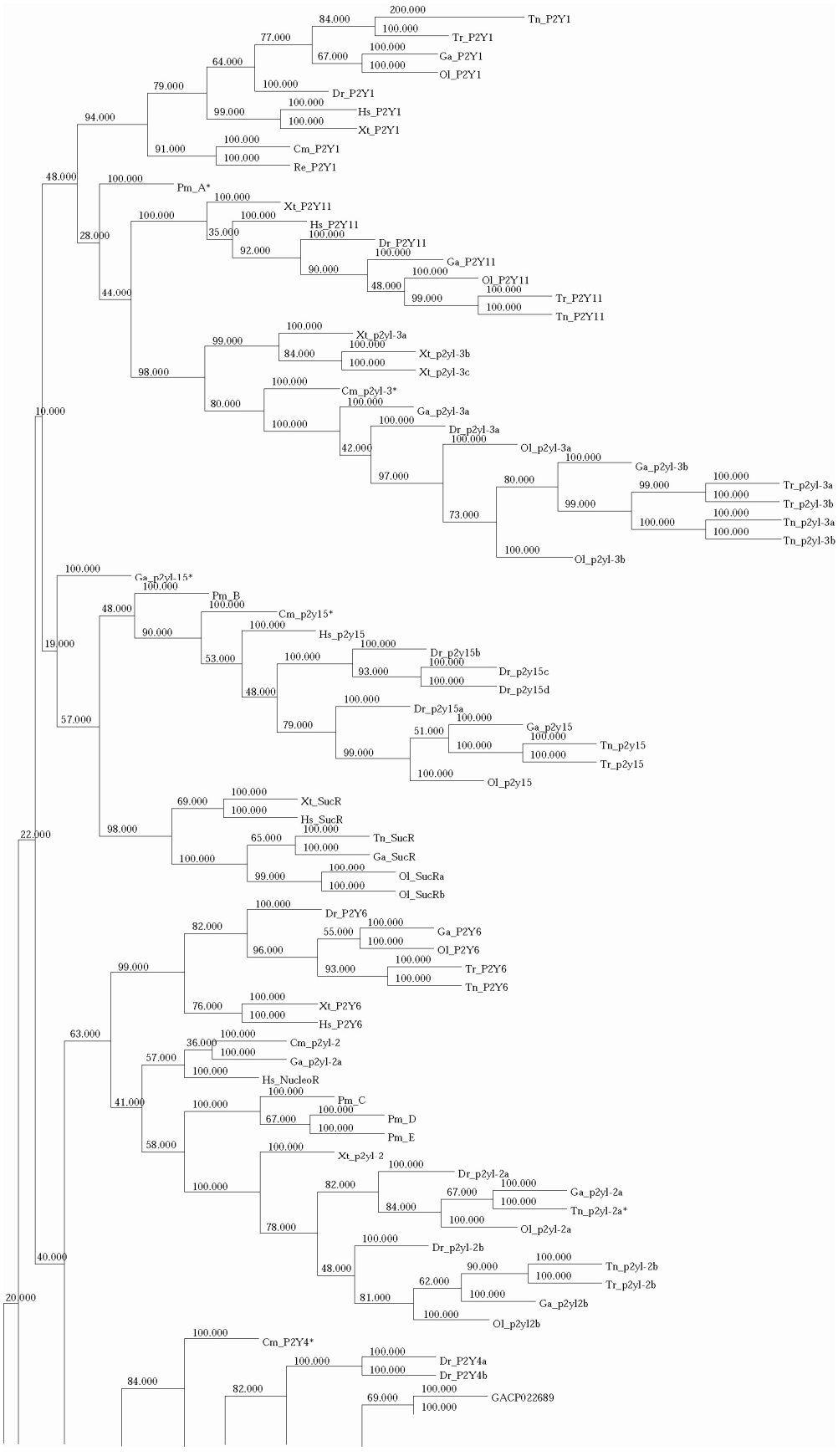
7.1 List of abbreviations

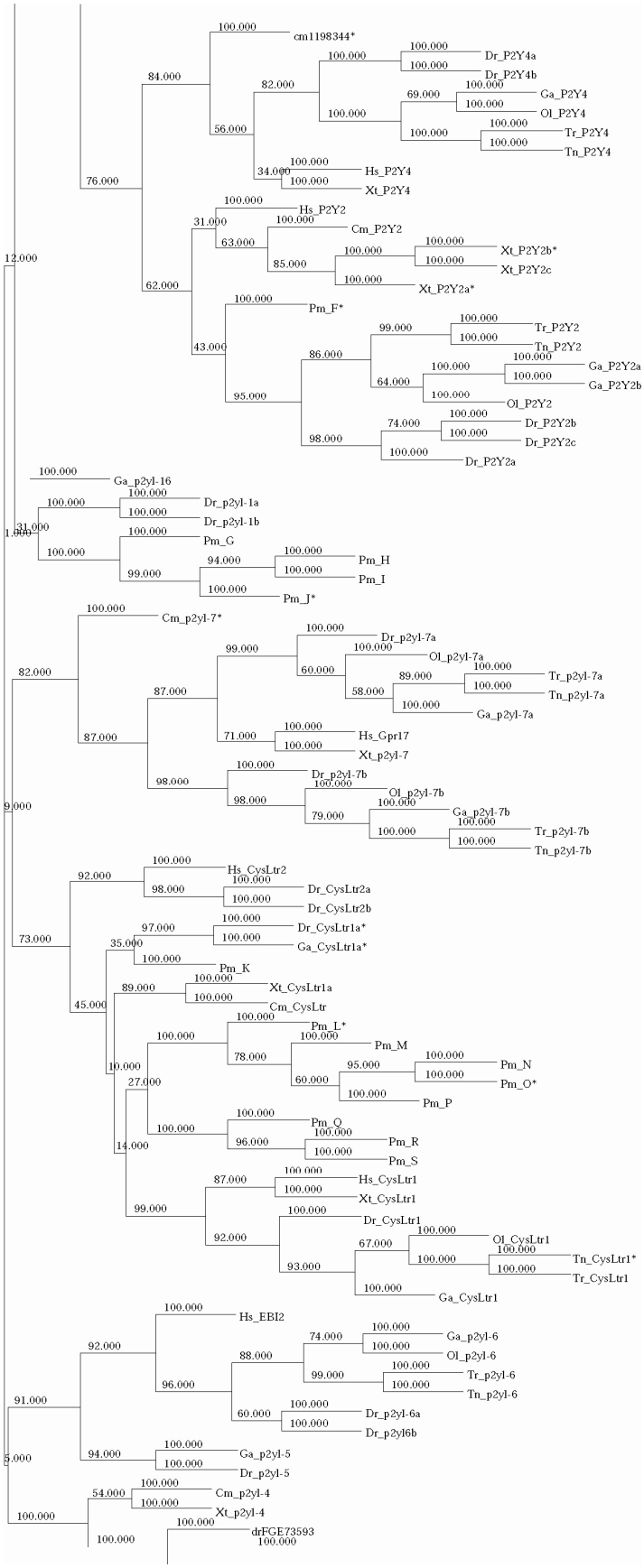
A, adenosine
ADP, adenosine diphosphate
AMP, adenosine monophosphate
AP, alkaline phosphatase
ATP, adenosine triphosphate
bp, base pairs
BSA, bovine serum albumin
cAMP, cyclic adenosine monophosphate
cDNA, complementary DNA
CysLtr, cysteine leukotriene
DEPC, diethylpyrocarbonate
DIG, digoxigenin
dpf, days post fertilization
DNA, deoxyribonucleic acid
dNTP, deoxyribonucleic triphosphate
EC, ectracellular domain
EDTA, ethylenediaminetetraacetic acid
FFAR, free-fatty acid receptor
GPCR, G-protein coupled receptor
HEK 293, human embryonic kidney cells
IC, intracellular domain
IPTG, isopropyl-_-D-1-thiogalactopyranoside
kb, kilo base
M, molar
MCS, multiple cloning site
µg, microgram
min, minutes
ng, nanogram
OE, olfactory epithelium
OB, olfactory bulb
OMP, olfactory marker protein
OR, olfactory receptor
OSN, olfactory sensory neuron
PBS, phosphate buffered saline
PCR, polymerase chain reaction
PFA, paraformaldehyde
RNA, ribonucleic acid
TE, tris-EDTA
TM, transmembrane
U, unit
UDP, uridine diphosphate
UTP, uridine triphosphate
X-Gal, 5-Bromo-4-chlor-3-indoyl-β-D-galactopyranoside

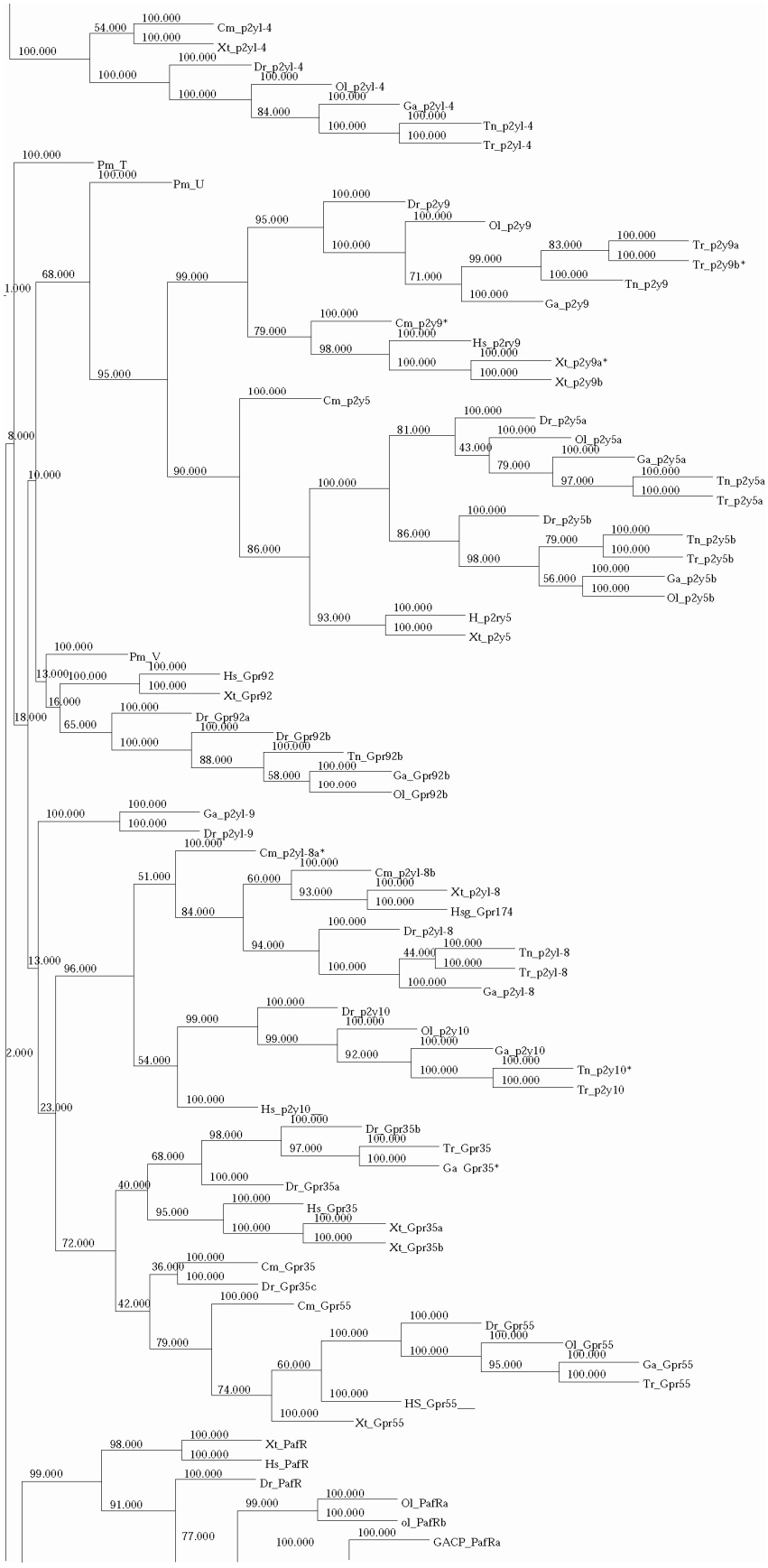
7.2 Supplemental figures

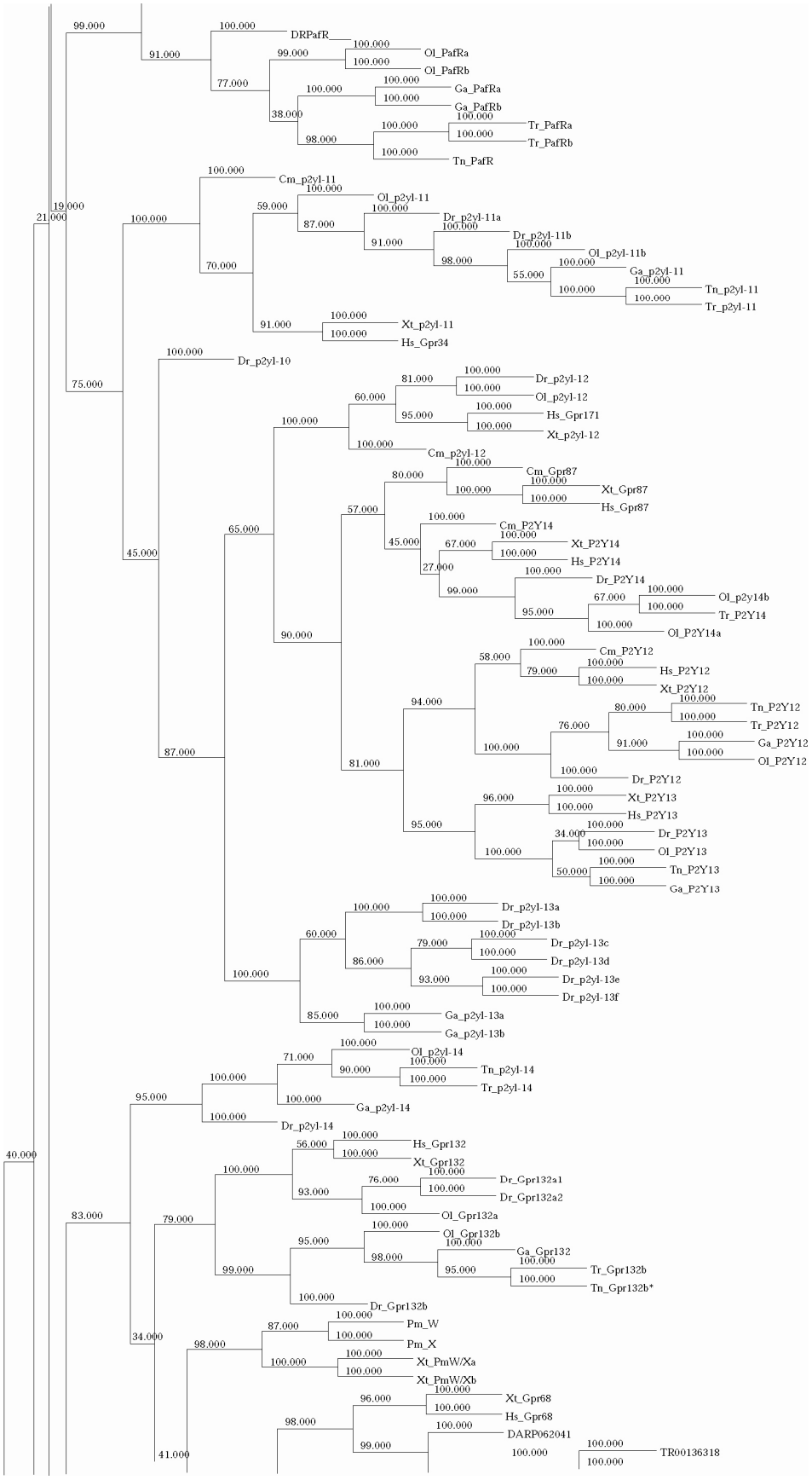
A. Phylogenetic tree of 8 fish species, frog and human

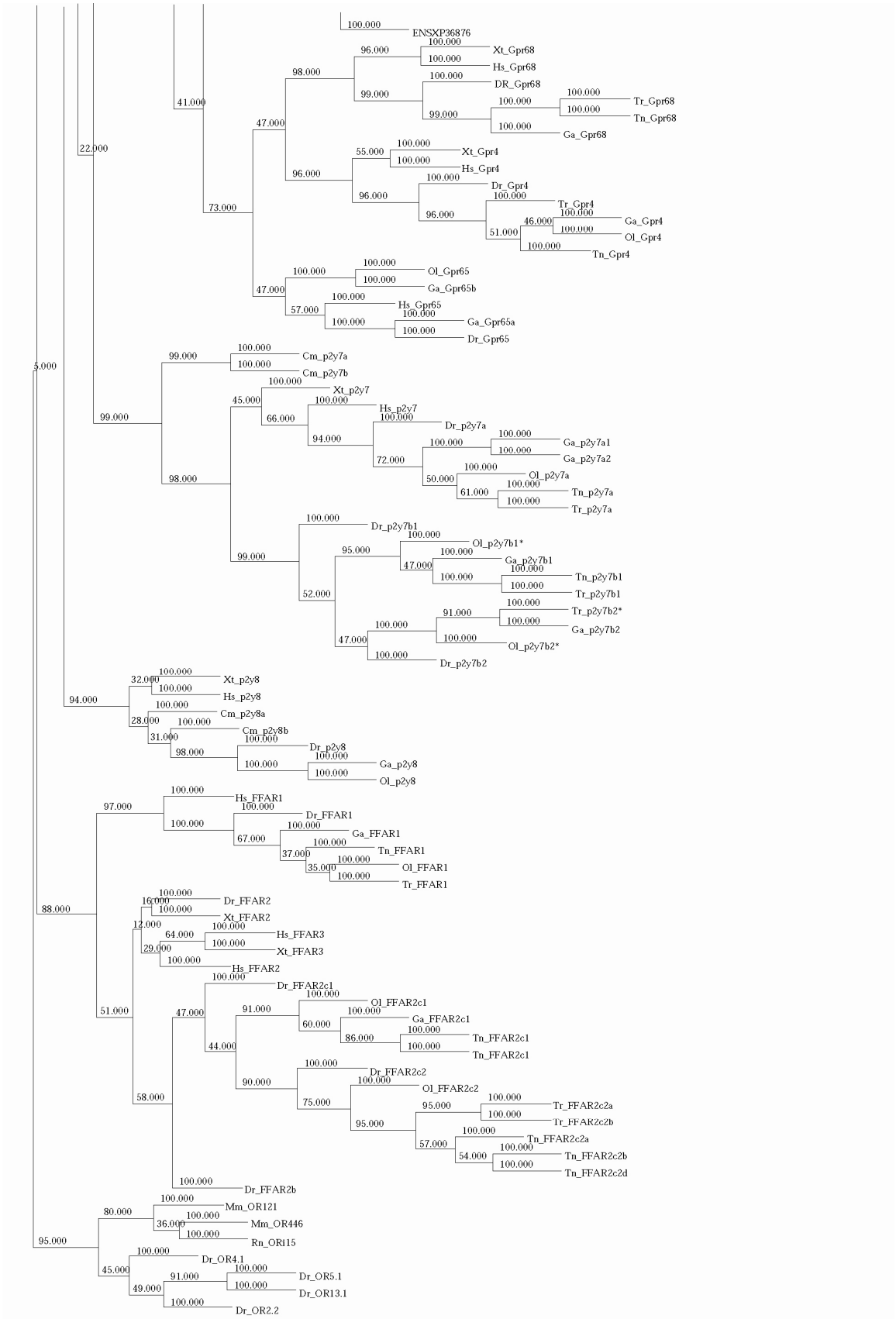
Abbreviations : Pm, lamprey; Cm, elephant shark, Re, little skate; Dr, zebrafish; Ga, stickleback; Ol, medaka; Tn, tetraodon; Tr, fugu; Xt, frog; Hs, human; PafR, platelet-activating receptor; SucR, succinate receptor; Gpr, G-protein coupled receptor; CysLtr, cysteine leukotriene; OR, olfactory receptor


Tn_P2Y1
Tr_P2Y1
Ga_P2Y1
Ol_P2Y1
Dr_P2Y1
Hs_P2Y1
Xt_P2Y1
Cm_P2Y1
Re_P2Y1
Pm_A*
Xt_P2Y11
Hs_P2Y11
Dr_P2Y11
Ga_P2Y11
Ol_P2Y11
Tr_P2Y11
Tn_P2Y11
Xt_p2yl-3a
Xt_p2yl-3b
Xt_p2yl-3c
Cm_p2yl-3*
Ga_p2yl-3a
Dr_p2yl-3a
Ol_p2yl-3a
Ga_p2yl-3b
Tr_p2yl-3a
Tr_p2yl-3b
Tn_p2yl-3a
Tn_p2yl-3b
Ol_p2yl-3b
Ga_p2yl-15*
Pm_B
Cm_p2y15*
Hs_p2y15
Dr_p2y15b
Dr_p2y15c
Dr_p2y15d
Dr_p2y15a
Ga_p2y15
Tn_p2y15
Tr_p2y15
Ol_p2y15
Xt_SucR
Hs_SucR
Tn_SucR
Ga_SucR
Ol_SucRa
Ol_SucRb
Dr_P2Y6
Ga_P2Y6
Ol_P2Y6
Tr_P2Y6
Tn_P2Y6
Xt_P2Y6
Hs_P2Y6
Cm_p2yl-2
Ga_p2yl-2a
Hs_NucleoR
Pm_C
Pm_D
Pm_E
Xt_p2yl-2
Dr_p2yl-2a
Ga_p2yl-2a
Tn_p2yl-2a*
Ol_p2yl-2a
Dr_p2yl-2b
Tn_p2yl-2b
Tr_p2yl-2b
Ga_p2yl2b
Ol_p2yl2b
Cm_P2Y4*
Dr_P2Y4a
Dr_P2Y4b
GACP022689


cm1198344*
Dr_P2Y4a
Dr_P2Y4b
Ga_P2Y4
Ol_P2Y4
Tr_P2Y4
Tn_P2Y4
Hs_P2Y4
Xt_P2Y4
Hs_P2Y2
Cm_P2Y2
Xt_P2Y2b*
Xt_P2Y2c
Xt_P2Y2a*
Pm_F*
Tr_P2Y2
Tn_P2Y2
Ga_P2Y2a
Ga_P2Y2b
Ol_P2Y2
Dr_P2Y2b
Dr_P2Y2c
Dr_P2Y2a
Ga_p2yl-16
Dr_p2yl-1a
Dr_p2yl-1b
Pm_G
Pm_H
Pm_I
Pm_J*
Cm_p2yl-7*
Dr_p2yl-7a
Ol_p2yl-7a
Tr_p2yl-7a
Tn_p2yl-7a
Ga_p2yl-7a
Hs_Gpr17
Xt_p2yl-7
Dr_p2yl-7b
Ol_p2yl-7b
Ga_p2yl-7b
Tr_p2yl-7b
Tn_p2yl-7b
Hs_CysLtr2
Dr_CysLtr2a
Dr_CysLtr2b
Dr_CysLtr1a*
Ga_CysLtr1a*
Pm_K
Xt_CysLtr1a
Cm_CysLtr
Pm_L*
Pm_M
Pm_N
Pm_O*
Pm_P
Pm_Q
Pm_R
Pm_S
Hs_CysLtr1
Xt_CysLtr1
Dr_CysLtr1
Ol_CysLtr1
Tn_CysLtr1*
Tr_CysLtr1
Ga_CysLtr1
Hs_EBI2
Ga_p2yl-6
Ol_p2yl-6
Tr_p2yl-6
Tn_p2yl-6
Dr_p2yl-6a
Dr_p2yl6b
Ga_p2yl-5
Dr_p2yl-5
Cm_p2yl-4
Xt_p2yl-4
drFGE73593


Cm_p2yl-4
Xt_p2yl-4
Dr_p2yl-4
Ol_p2yl-4
Ga_p2yl-4
Tn_p2yl-4
Tr_p2yl-4
Pm_T
Pm_U
Dr_p2y9
Ol_p2y9
Tr_p2y9a
Tr_p2y9b*
Tn_p2y9
Ga_p2y9
Cm_p2y9*
Hs_p2ry9
Xt_p2y9a*
Xt_p2y9b
Cm_p2y5
Dr_p2y5a
Ol_p2y5a
Ga_p2y5a
Tn_p2y5a
Tr_p2y5a
Dr_p2y5b
Tn_p2y5b
Tr_p2y5b
Ga_p2y5b
Ol_p2y5b
H_p2ry5
Xt_p2y5
Pm_V
Hs_Gpr92
Xt_Gpr92
Dr_Gpr92a
Dr_Gpr92b
Tn_Gpr92b
Ga_Gpr92b
Ol_Gpr92b
Ga_p2yl-9
Dr_p2yl-9
Cm_p2yl-8a*
Cm_p2yl-8b
Xt_p2yl-8
Hsg_Gpr174
Dr_p2yl-8
Tn_p2yl-8
Tr_p2yl-8
Ga_p2yl-8
Dr_p2y10
Ol_p2y10
Ga_p2y10
Tn_p2y10*
Tr_p2y10
Hs_p2y10
Dr_Gpr35b
Tr_Gpr35
Ga_Gpr35*
Dr_Gpr35a
Hs_Gpr35
Xt_Gpr35a
Xt_Gpr35b
Cm_Gpr35
Dr_Gpr35c
Cm_Gpr55
Dr_Gpr55
Ol_Gpr55
Ga_Gpr55
Tr_Gpr55
HS_Gpr55___
Xt_Gpr55
Xt_PafR
Hs_PafR
Dr_PafR
Ol_PafRa
ol_PafRb
GACP_PafRa


DRPafR
Ol_PafRa
Ol_PafRb
Ga_PafRa
Ga_PafRb
Tr_PafRa
Tr_PafRb
Tn_PafR
Cm_p2yl-11
Ol_p2yl-11
Dr_p2yl-11a
Dr_p2yl-11b
Ol_p2yl-11b
Ga_p2yl-11
Tn_p2yl-11
Tr_p2yl-11
Xt_p2yl-11
Hs_Gpr34
Dr_p2yl-10
Dr_p2yl-12
Ol_p2yl-12
Hs_Gpr171
Xt_p2yl-12
Cm_p2yl-12
Cm_Gpr87
Xt_Gpr87
Hs_Gpr87
Cm_P2Y14
Xt_P2Y14
Hs_P2Y14
Dr_P2Y14
Ol_p2y14b
Tr_P2Y14
Ol_P2Y14a
Cm_P2Y12
Hs_P2Y12
Xt_P2Y12
Tn_P2Y12
Tr_P2Y12
Ga_P2Y12
Ol_P2Y12
Dr_P2Y12
Xt_P2Y13
Hs_P2Y13
Dr_P2Y13
Ol_P2Y13
Tn_P2Y13
Ga_P2Y13
Dr_p2yl-13a
Dr_p2yl-13b
Dr_p2yl-13c
Dr_p2yl-13d
Dr_p2yl-13e
Dr_p2yl-13f
Ga_p2yl-13a
Ga_p2yl-13b
Ol_p2yl-14
Tn_p2yl-14
Tr_p2yl-14
Ga_p2yl-14
Dr_p2yl-14
Hs_Gpr132
Xt_Gpr132
Dr_Gpr132a1
Dr_Gpr132a2
Ol_Gpr132a
Ol_Gpr132b
Ga_Gpr132
Tr_Gpr132b
Tn_Gpr132b*
Dr_Gpr132b
Pm_W
Pm_X
Xt_PmW/Xa
Xt_PmW/Xb
Xt_Gpr68
Hs_Gpr68
DARP062041
TR00136318


ENSXP36876
Xt_Gpr68
Hs_Gpr68
DR_Gpr68
Tr_Gpr68
Tn_Gpr68
Ga_Gpr68
Xt_Gpr4
Hs_Gpr4
Dr_Gpr4
Tr_Gpr4
Ga_Gpr4
Ol_Gpr4
Tn_Gpr4
Ol_Gpr65
Ga_Gpr65b
Hs_Gpr65
Ga_Gpr65a
Dr_Gpr65
Cm_p2y7a
Cm_p2y7b
Xt_p2y7
Hs_p2y7
Dr_p2y7a
Ga_p2y7a1
Ga_p2y7a2
Ol_p2y7a
Tn_p2y7a
Tr_p2y7a
Dr_p2y7b1
Ol_p2y7b1*
Ga_p2y7b1
Tn_p2y7b1
Tr_p2y7b1
Tr_p2y7b2*
Ga_p2y7b2
Ol_p2y7b2*
Dr_p2y7b2
Xt_p2y8
Hs_p2y8
Cm_p2y8a
Cm_p2y8b
Dr_p2y8
Ga_p2y8
Ol_p2y8
Hs_FFAR1
Dr_FFAR1
Ga_FFAR1
Tn_FFAR1
Ol_FFAR1
Tr_FFAR1
Dr_FFAR2
Xt_FFAR2
Hs_FFAR3
Xt_FFAR3
Hs_FFAR2
Dr_FFAR2c1
Ol_FFAR2c1
Ga_FFAR2c1
Tn_FFAR2c1
Tn_FFAR2c1
Dr_FFAR2c2
Ol_FFAR2c2
Tr_FFAR2c2a
Tr_FFAR2c2b
Tn_FFAR2c2a
Tn_FFAR2c2b
Tn_FFAR2c2d
Dr_FFAR2b
Mm_OR121
Mm_OR446
Rn_OR115
Dr_OR4.1
Dr_OR5.1
Dr_OR13.1
Dr_OR2.2

B. P2Y genes accession number and/or IDs, and location

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Zebrafish		ENSDARP00000004774	3	53,541,140-53,542,045.
Zebrafish	NP_955900.1	ENSDARP00000032650	2	43,079,372-43,080,946.
Zebrafish	XP_001335235.1	ENSDARP00000045017	14	37,018,669-37,019,712.
Zebrafish	XP_001343519.1	ENSDARP00000058877	1	879,202-880,541.
Zebrafish	XP_694459.1	ENSDARP00000062479	18	25,785,514-25,787,196.
Zebrafish	XP_001341296.1	FGENESH00000078739	24	10,123,808-10,124,984.
Zebrafish		ENSDARP00000070013	14	37,146,462-37,147,913.
Zebrafish		ENSDARP00000075538	14	6,365,702-6,366,698.
Zebrafish		GENSCAN00000023739	5	21,317,494
Zebrafish		ENSDARP00000081437	5	21,321,347-21,322,297
Zebrafish		FGENESH00000078957	13	50,798,456-50,842,878.
Zebrafish	XP_001336913.1	FGENESH00000058659	Zv7_NA1868	12,840-19,467.
Zebrafish		ENSDARP00000087296	Zv7_NA1868	12,636-33,632.
Zebrafish	XP_001336698.1	FGENESH00000071328	1	893,686-894,675
Zebrafish	XP_001339996.1	ENSDARP00000087397	Scaffold Zv7_NA791	33,990-34,949.
Zebrafish	XP_684762.1	ENSDARP00000010296	9	25,044,989-25,046,056.
Zebrafish	NP_001035473.1	ENSDARP00000041935	1	903,492-907,044.
Zebrafish	XP_001342443.1	ENSDARP00000052979	7	17,039,373-17,040,398.
Zebrafish	XP_001346142.1	FGENESH00000084837	23	24,448,733-24,449,707.
Zebrafish		ENSDARP00000060098	9	25,074,319-25,075,413
Zebrafish	NP_001035080.1	FGENESH00000073641	15	35,485,427-35,486,407.
Zebrafish	XP_693153.1	FGENESH00000073593	6	45,082,130-45,082,981.
Zebrafish	XP_001333765.1	ENSDARP00000069269	2	43,060,465-43,062,121.
Zebrafish	XP_001334713.1	ENSDARP00000075542	14	6,395,616-6,396,587.
Zebrafish	XP_001343355.1	ENSDARP00000080918	18	44,456,810-44,457,877
Zebrafish	XP_687425.2	ENSDARP00000083420	6	19,484,326-19,485,273.
Zebrafish	NP_001073524.1	ENSDARP00000085057	23	13,827,907-13,834,737.
Zebrafish	AAI34132.1	FGENESH00000061740	14	6,408,647-6,419,175.
Zebrafish	XP_700330.1	FGENESH00000066244	18	1,637,660-1,638,214
Zebrafish	XP_691569.2	FGENESH00000071330	15	42,775,008-42,777,576.
Zebrafish	XP_697876.2	ENSDARP00000044742	9	23,293,650-23,294,480.
Zebrafish	XP_001342372.1	ENSDARP00000062129	17	33,899,631-33,901,492.
Zebrafish		ENSDARP00000068982	1	33,627,665-33,631,029.
Zebrafish	XP_001332886.1	FGENESH00000058248	10	13,607,347-13,608,423.
Zebrafish		GENSCAN00000005772	5	21,317,594-21,318,487
Zebrafish		FGENESH00000070496	7	17,426,371-17,428,125.
Zebrafish		FGENESH00000061858	14	6,499,904-6,500,887
Zebrafish	XP_001340065.1	ENSDARP00000087396	18	1,659202-1,660,173
Zebrafish	XP_692215.1	FGENESH00000051734	15	25,619,485-25,620,579
Zebrafish		ENSDARP00000062041	17	37,840,794-37,841,684.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Zebrafish		ENSDARP00000084735	15	42,665,690-42,690,969.
Zebrafish		ENSDARP00000067542	17	2,870,988-2,872,004.
Zebrafish		ENSDARP00000030390	17	38,878,390-38,879,292.
Zebrafish	NM_001025547.1	ENSDARP00000063752	20	18,442,740-18,449,712.
Zebrafish	XP_686786.2	FGENESH00000066881	5	34,052,650-34,054,470.
Zebrafish		ENSDARP00000072403	9	28,958,014-28,960,102.
Zebrafish		ENSDARP00000079157	14	41,893,740-41,894,729
Zebrafish		ENSDARP00000086767	Zv7_NA117	
Zebrafish	XP_001342441.1	GENSCAN00000008806	6	39,411,564-39,412,598.
Zebrafish	NP_001076364.1	GENSCAN00000015928	19	8,175,835-8,176,635.
Zebrafish	NP_898896.1	ENSDARP00000007507	17	2,893,697-2,894,776.
Zebrafish		ENSDARP00000069332	Zv7_NA992	
Zebrafish		ENSDARP00000078231	16	50,292,613-50,293,762.
Zebrafish		ENSDARP00000075859	19	8,189,303-8,190,178.
Zebrafish		ENSDARP00000086328	19	8,194,384-8,195,385
Zebrafish	XP_001338686.1	FGENESH00000073598	15	35,474,314-35,475,153.
Zebrafish	XP_001343968.1	FGENESH00000055237	21	9,205,389-9,206,556.
Zebrafish	XP_001334570.1	ENSDARP00000009604	2	20,384,476-20,385,552.
Zebrafish		ENSDARP00000009884	7	24,766,695-24,767,759.
Zebrafish	XP_001337554.1	FGENESH00000082569.	24	4,108,722-4,109,708.
Zebrafish	NP_001013368.1	ENSDARP00000051531	5	12,784,540-12,791,594.
Zebrafish	NP_001025368.1	GENSCAN00000041734	8	51,058,965-51,060,442.
Zebrafish	XP_001330993.1	ENSDARP00000027382	9	27,386,792-27,388,148.
Zebrafish		ENSDARP00000022016	4	39,121,691-39,122,806.
Zebrafish		ENSDARP00000055274	6	59,132,981-59,134,152.
Zebrafish	XP_001340906.1	GENSCAN00000038826	9	17,394,200-17,395,457.
Zebrafish	ABI99470.1	FGENESH00000070088	8	51,058,965-51,060,442.
Zebrafish	XP_001339169.1	FGENESH00000062132	3	41,370,113-41,377,705.
Zebrafish	XP_001341328.1	ENSDARP00000025567	12	33,919,188-33,922,860.
Zebrafish	XP_694953.1	ENSDARP00000050272	22	24,654,997-24,665,360.
Zebrafish		ENSDARP00000011684	15	4,357,742-4,362,018.
Zebrafish		ENSDARP00000076026	15	4,461,247-4,462,197
Zebrafish		ENSDARP00000021049	15	4,413,959-4,415,307.
Zebrafish		ENSDARP00000087292	15	4,402,047-4,403,019.
Zebrafish		ENSDARP00000087308	15	4,374,155-4,374,790
Zebrafish		ENSDARP00000062905	15	4,451,917-4,452,690
Zebrafish		ENSDARP00000087287	15	4,421,132-4,423,731.
Zebrafish		ENSDARP00000087299	15	4,387,262-4,388,179
Zebrafish		ENSDARP00000091856	21	18,992,364-18,994,367.
Zebrafish		ENSDARP00000092997	15	3,026,680-3,030,111.
Zebrafish		ENSDARP00000093000	15	3,010,012-3,012,469.
Zebrafish		ENSDARP00000057319	13	36,681,559-36,683,336.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Zebrafish		ENSDARP00000088649	22	24,613,687-24,617,923.
Zebrafish		ENSDARP00000070796	2	34,986,152-34,990,484.
Stickleback		ENSGACP00000000942	scaffold_47	184,282-185,445.
Stickleback		ENSGACP00000002502	scaffold_132	209,478-211,560.
Stickleback		ENSGACP00000004999	groupVIII	1,931,956-1,933,858.
Stickleback		ENSGACP00000007390	groupI	1,730,484-1,731,398.
Stickleback		ENSGACP00000010855	groupI	6,840,543-6,841,996.
Stickleback		ENSGACP00000012761	groupXV	7,895,097-7,896,633.
Stickleback		ENSGACP00000013551	groupI	10,467,941-10,468,882.
Stickleback		ENSGACP00000015144	groupXVII	12,409,961-12,410,890.
Stickleback		ENSGACP00000018133	groupXI	13,421,555-13,422,436.
Stickleback		ENSGACP00000022689	groupIV	5,767,958-5,769,001.
Stickleback		ENSGACP00000024554	groupIV	15,741,075-15,742,240.
Stickleback		ENSGACP00000026536	groupVII	10,390,881-10,392,102.
Stickleback		ENSGACP00000002451	groupXVI	2,783,893-2,785,408.
Stickleback		ENSGACP00000003378	scaffold_475	5,558-8,755.
Stickleback		ENSGACP00000005028	groupXXI	8,709,829-8,711,970.
Stickleback		ENSGACP00000007789	groupXIX	7,281,088-7,282,050.
Stickleback		ENSGACP00000010882	groupI	6,917,158-6,918,464.
Stickleback		ENSGACP00000013482	groupI	10,309,588-10,310,505.
Stickleback		ENSGACP00000014878	groupXII	15,037,417-15,039,097.
Stickleback		ENSGACP00000015806	groupI	14,341,623-14,343,831.
Stickleback		ENSGACP00000018449	groupI	20,645,156-20,648,034.
Stickleback		ENSGACP00000024549	groupIV	15,726,274-15,727,675.
Stickleback		ENSGACP00000026938	groupVII	17,758,768-17,760,612.
Stickleback		ENSGACP00000002454	groupXVI	2,792,172-2,793,433.
Stickleback		ENSGACP00000005276	groupXIX	5,486,695-5,487,630.
Stickleback		ENSGACP00000010887	groupI	6,926,441-6,928,012.
Stickleback		ENSGACP00000013543	groupI	10,382,050-10,383,467.
Stickleback		ENSGACP00000015138	groupXII	15,233,419-15,235,543.
Stickleback		ENSGACP00000017795	groupIII	1,504,743-1,505,922.
Stickleback		ENSGACP00000020237	groupIII	7,582,033-7,583,389.
Stickleback		ENSGACP00000024551	groupIV	15,735,707-15,737,199.
Stickleback		ENSGACP00000024559	groupIV	15,772,138-15,773,109.
Stickleback		ENSGACP00000002511	scaffold_128	71,149-75,065.
Stickleback		ENSGACP00000013980	groupXV	9,528,005-9,531,950.
Stickleback		ENSGACP00000026368	groupVII	9,005,340-9,006,568.
Stickleback		ENSGACP00000013981	groupXV	9,533,782-9,534,439.
Stickleback		ENSGACP00000012423	groupI	9,090,447-9,091,400.
Stickleback		ENSGACP00000016705	groupI	16,310,314-16,314,973.
Stickleback		ENSGACP00000002567	groupXVI	3,208,336-3,211,795.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Stickleback		ENSGACP00000013383	groupXIII	9,878,841-9,880,899.
Stickleback		ENSGACP00000020793	groupXIV	856,996-858,082.
Stickleback		ENSGACP00000020791	groupXIV	844,787-846,001.
Stickleback		ENSGACP00000021482	groupIV	94,613-97,364.
Stickleback		ENSGACP00000013380	groupXIII	9,865,275-9,867,291.
Stickleback		ENSGACP00000020792	groupXIV	851,632-855,772.
Stickleback		ENSGACP00000013995	groupXV	9,558,404-9,560,352.
Stickleback		ENSGACP00000003220	groupX	2,121,801-2,122,745.
Stickleback		ENSGACP00000003223	groupX	2,125,535-2,126,404
Stickleback		ENSGACP00000004815	groupXXI	8,276,903-8,278,051.
Stickleback		ENSGACP00000013986	groupXV	9,548,590-9,550,771.
Stickleback		ENSGACP00000024507	groupIV	15,518,375-15,519,559.
Stickleback		ENSGACP00000005053	groupXIX	5,063,430-5,064,581.
Stickleback		ENSGACP00000013989	groupXV	9,550,633-9,554,068.
Stickleback		ENSGACP00000013985	groupXV	9,544,137-9,545,953.
Stickleback		ENSGACP00000013991	groupXV	9,554,340-9,556,929.
Stickleback		ENSGACP00000010896	groupXVI	16,119,568-16,120,711.
Stickleback		ENSGACP00000019865	groupI	24,965,025-24,966,188
Stickleback		ENSGACP00000006492	groupI	628,832-630,104.
Stickleback		ENSGACP00000013505	groupI	10,334,339-10,336,025.
Stickleback		ENSGACP00000006495	groupI	638,917-644,877.
Stickleback		ENSGACP00000021871	groupIX	4,534,658-4,536,191.
Stickleback		ENSGACP00000013498	groupI	10,325,442-10,326,592.
Stickleback		ENSGACP00000018857	groupIII	3,261,595-3,267,024.
Stickleback		ENSGACP00000025060	groupVII	1,825,852-1,827,570.
Stickleback		ENSGACP00000025070	groupVII	1,847,986-1,849,223.
Stickleback		ENSGACP00000025072	groupVII	1,852,514-1,854,003.
Stickleback		ENSGACP00000023536	groupIII	16,033,511-16,034,929.
Medaka		ENSORLP00000002219	10	7,033,469-7,035,501.
Medaka		GENSCAN00000054265	scaffold6_contig5754	3,835-4,848.
Medaka		ENSORLP00000007398	13	12,276,545-12,277,609.
Medaka		ENSORLP00000012926	10	23,776,892-23,778,001.
Medaka		GENSCAN00000058831	scaffold40_contig27481	1,322-2,938.
Medaka		ENSORLP00000013995	21	9,360,855-9,362,297.
Medaka		GENSCAN00000093228	scaffold113_contig52046	870-2,045.
Medaka		ENSORLP00000016967	14	24,582,440-24,585,642.
Medaka		ENSORLP00000019532	13	32,065,644-32,068,080.
Medaka		ENSORLP00000020401	7	25,577,394-25,579,349.
Medaka		ENSORLP00000020821	17	25,697,062-25,699,648.
Medaka		ENSORLP00000003758	13	6,303,968-6,307,443.
Medaka		ENSORLP00000007457	13	12,413,570-12,414,881.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Medaka		ENSORLP00000010446	14	16,429,346-16,430,576.
Medaka		ENSORLP00000012929	20	17,310,797-17,311,732.
Medaka		GENSCAN00000039365	scaffold31_contig22747	865-1,788.
Medaka		ENSORLP00000007471	13	12,420,330-12,422,728.
Medaka		ENSORLP00000012907	13	19,920,732-19,923,609.
Medaka		GENSCAN00000122403	scaffold124_contig54616	1,369-2,403.
Medaka		ENSORLP00000012933	10	23,791,767-23,800,457.
Medaka		ENSORLP00000012975	18	25,499,030-25,500,316.
Medaka		GENSCAN00000118089	scaffold87_contig44759	1,479-2,606.
Medaka		GENSCAN00000078942.	scaffold113_contig52045	1,818-3,814.
Medaka		GENSCAN00000043172	scaffold113_contig52044	19,805-22,435.
Medaka		ENSORLP00000019495	4	28,526,262-28,528,623.
Medaka		GENSCAN00000094078	scaffold12_contig11312	12,331-13,296.
Medaka		GENSCAN00000043609	scaffold265_contig80368	1,608-2,663.
Medaka		ENSORLP00000012595	20	16,658,187-16,659,464.
Medaka		ENSORLP00000001770	13	2,971,912-2,978,149.
Medaka		ENSORLP00000008955	13	14,370,077-14,371,428.
Medaka		ENSORLP00000004359	18	4,855,220-4,857,861.
Medaka		ENSORLP00000016190	22	6,473,385-6,476,684.
Medaka		ENSORLP00000022742	24	23,947,426-23,948,451.
Medaka		ENSORLP00000002716	12	2,487,239-2,488,102.
Medaka		ENSORLP00000013679	21	8,754,070-8,758,495.
Medaka		ENSORLP00000002719	12	2,500,854-2,505,755.
Medaka		ENSORLP00000012243	9	13,758,433-13,759,293.
Medaka		ENSORLP00000002728	12	2,510,670-2,512,451.
Medaka		ENSORLP00000012252	9	13,773,407-13,775,175.
Medaka		ENSORLP00000016151	22	6,420,792-6,424,168.
Medaka		ENSORLP00000016176	22	6,435,128-6,437,006.
Medaka		ENSORLP00000016172	22	6,430,001-6,431,425.
Medaka		ENSORLP00000011345	11	18,725,568-18,726,452.
Medaka		ENSORLP00000018838	16	22,038,177-22,039,559.
Medaka		ENSORLP00000011361	11	18,729,700-18,731,512.
Medaka		ENSORLP00000010534	10	18,866,733-18,867,974.
Medaka		ENSORLP00000016185	22	6,440,215-6,442,229.
Medaka		ENSORLP00000006057	2	28,193,626-28,194,765.
Medaka		ENSORLP00000017154	21	14,053,477-14,054,570.
Medaka		ENSORLP00000000417	13	188,698-189,741.
Medaka		ENSORLP00000000420	13	200,515-202,585.
Medaka		ENSORLP00000007515	13	12,492,576-12,493,881.
Medaka		ENSORLP00000000443	13	226,304-229,283.
Medaka		ENSORLP00000007530	13	12,502,909-12,504,826.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Medaka		ENSORLP00000012908	13	19,921,273-19,924,973.
Medaka		ENSORLP00000015583	17	18,607,766-18,617,315.
Medaka		ENSORLP00000023635	scaffold474	194,770-197,848.
Medaka		ENSORLP00000023642	scaffold474	207,648-208,411.
Medaka		ENSORLP00000003694	17	3,206,894-3,208,153.
Tetraodon	CAF90112.1	GSTENP00004207001	1	9,280,563-9,281,691.
Tetraodon	CAF97358.1	GSTENP00014693001	7	9,860,643-9,862,235.
Tetraodon	CAG00913.1	GSTENP00019495001	10	9,464,265-9,465,341.
Tetraodon	CAG03809.1	GSTENP00023354001	9	10,281,976-10,282,794.
Tetraodon	CAG04290.1	GSTENP00023967001	6	3,782,208-3,782,984.
Tetraodon		GSTENP00026760001	16	7,549,182-7,557,621.
Tetraodon	CAG10424.1	GSTENP00032057001	10	7,304,828-7,305,844.
Tetraodon	CAG14539.1	GSTENP00036486001	1	9,273,208-9,273,996.
Tetraodon		GSTENP00004208001	1	9,295,942-9,296,815.
Tetraodon	CAF97378.1	GSTENP00014719001	16	2,404,478-2,405,551.
Tetraodon	CAG00917.1	GSTENP00019499001	10	9,527,289-9,528,587.
Tetraodon	CAG03842.1	GSCT00003286001	9	10,082,246-10,086,911.
Tetraodon	CAG04953.1	GSTENP00024817001	3	2,951,632-2,952,531.
Tetraodon	CAG07299.1	GSTENP00027920001	2	7,883,721-7,884,698.
Tetraodon	CAG12186.1	GSTENP00034382001	3	11,549,843-11,550,697.
Tetraodon	CAG14148.1	GSTENP00037810001	1	9,270,565-9,271,759.
Tetraodon		GSTENP00009815001	Un_random	116,432,397-116,448,838
Tetraodon		GSTENP00014725001	16	2,343,694-2,344,584.
Tetraodon	CAG03844.1	GSTENP00023395001	9	10,073,298-10,074,278.
Tetraodon	CAG05650.1	GSTENP00025746001	1_random	735,645-736,658.
Tetraodon	CAG09809.1	GSTENP00031240001	1	8,499,289-8,500,194.
Tetraodon	CAG12785.1	GSTENP00035226001	16	5,864,119-5,865,153.
Tetraodon		GSTENP00014840001	16	1,322,744-1,323,541.
Tetraodon	CAG12733.1	GSTENP00035167001	16	6,377,491-6,378,788.
Tetraodon		GSTENP00016932001	10	1,794,376-1,796,696.
Tetraodon		GSTENP00035124001	Un_random	20,462,662-20,464,181.
Tetraodon	CAG02919.1	GSTENP00022162001	7	5,193,542-5,194,447.
Tetraodon		GSTENP00028691001	12	2,774,950-2,775,867.
Tetraodon		GSTENP00028689001	12	2,785,371-2,786,216.
Tetraodon		GSTENP00025234001	Un_random	1,558,254-1,559,111.
Tetraodon		GSTENP00025233001	Un_random	1,563,434-1,567,759.
Tetraodon		GSTENP00025232001	Un_random	1,569,345-1,570,370.
Tetraodon		GSTENG00012077001	Un_random	47,547,810-47,548,679.
Tetraodon		GSTENP00001382001	Un_random	71,283,807-71,284,259.
Tetraodon		GSTENP00032031001	10	7,479,304-7,483,281.
Tetraodon	CAF89689.1	GSTENP00003669001	Un_random	83,331,664-83,332,485.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Tetraodon		GSTENP00022723001	Un_random	22,471,163-22,472,059.
Tetraodon		GSTENP00022725001	Un_random	22,483,850-22,484,861.
Tetraodon		GSTENP00001856001	Un_random	57,930,073-57,930,654.
Tetraodon		GSTENP00003668001	Un_random	83,327,996-83,329,049.
Tetraodon		GSTENP00008852001	Un_random	110,447,989-110,448,855.
Tetraodon	CAG04270.1	GSTENP00023936001	6	3,386,437-3,387,465.
Tetraodon		GSTENP00029336001	1	4,022,343-4,023,317
Tetraodon		GSTENP00027186001	3	5,654,006-5,655,053.
Tetraodon	CAF94854.1	GSTENP00011245001	Un_random	124,740,695-124,741,660.
Tetraodon		GSTENP00014732001	16	2,302,536-2,303,345.
Tetraodon		GSTENP00014733001	16	2,250,394-2,251,875.
Tetraodon		GSTENP00010007001	Un_random	117,578,407-117,580,225.
Tetraodon		GSTENP00014269001	Un_random	21,987,145-21,988,507.
Tetraodon		GSTENP00014270001	Un_random	21,984,004-21,984,779.
Fugu		SINFRUP00000141264	scaffold_417	98,990-100,638.
Fugu		GENSCANSLICE00000028178	scaffold_163	452,712-453,946.
Fugu		SINFRUP00000159316	scaffold_36	699,699-700,589.
Fugu		SINFRUP00000179409	scaffold_19	1,437,922-1,438,965.
Fugu		SINFRUP00000181180	scaffold_15	233,344-234,201.
Fugu		SINFRUP00000182177	scaffold_132	490,310-491,281.
Fugu		SINFRUP00000141960	scaffold_6	1,757,599-1,762,074.
Fugu		SINFRUP00000151445	scaffold_173	468,694-469,756.
Fugu		SINFRUP00000161432	scaffold_31	471,544-481,436.
Fugu		SINFRUP00000177522	scaffold_13	126,997-128,013.
Fugu		SINFRUP00000179625	scaffold_344	182,644-183,670.
Fugu		SINFRUP00000181229	scaffold_10	1,241,330-1,242,346.
Fugu		SINFRUP00000183021	scaffold_9	155,401-156,462.
Fugu		SINFRUP00000147658	scaffold_142	539,835-540,680.
Fugu		SINFRUP00000151908	scaffold_200	69,219-70,007.
Fugu		SINFRUP00000165330	scaffold_173	403,846-405,443.
Fugu		SINFRUP00000179172	scaffold_304	292,446-293,420.
Fugu		SINFRUP00000180200	scaffold_163	433,438-434,301.
Fugu		SINFRUP00000181470	scaffold_2435	816-10,342.
Fugu		GENSCANSLICE00000028179	scaffold_163	457,415-459,054.
Fugu		GENSCANSLICE00000028180	scaffold_163	461,378-462,478.
Fugu		SINFRUP00000181476	scaffold_142	156,792-157,868.
Fugu		SINFRUP00000178171	scaffold_304	59,403-60,263.
Fugu		SINFRUP00000128368	scaffold_19	1,991,283-1,993,128.
Fugu		SINFRUP00000136318	scaffold_34	596,236-601,178.
Fugu		SINFRUP00000178764	scaffold_757	25,197-26,213.
Fugu		SINFRUP00000177569	scaffold_9	1,251,655-1,252,494.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Fugu		SINFRUP00000180629	scaffold_276	317,473-318,520.
Fugu		SINFRUP00000130845	scaffold_27	140,400-141,482.
Fugu		SINFRUP00000180210	scaffold_27	153,992-154,652.
Fugu		SINFRUP00000130847	scaffold_27	143,059-146,762.
Fugu		SINFRUP00000182548	scaffold_4	636,672-637,589.
Fugu		SINFRUP00000178146	scaffold_4	647,031-647,867.
Fugu		SINFRUP00000150183	scaffold_10	1,426,614-1,427,435.
Fugu		SINFRUP00000150187	scaffold_10	1,431,395-1,436,521.
Fugu		SINFRUP00000178352	scaffold_69	1,033,517-1,034,392.
Fugu		SINFRUP00000180383	scaffold_573	73,946-74,974.
Fugu		SINFRUP00000179870	scaffold_69	1,030,701-1,031,549.
Fugu		SINFRUP00000180345	scaffold_35	257,783-278,516.
Fugu		SINFRUP00000128201	scaffold_5	451,609-452,583.
Fugu		SINFRUP00000134030	scaffold_2	2,814,874-2,816,085.
Fugu	Q6XCB2_FUGRU	SINFRUP00000134141	scaffold_32	973,830-974,966.
Fugu		SINFRUP00000138083	scaffold_9	270,357-271,854.
Fugu		SINFRUP00000180272	scaffold_12867	1,217-2,206.
Fugu		SINFRUP00000166034	scaffold_2529	63-482.
Fugu		SINFRUP00000150750	scaffold_163	452,796-462,394.
Fugu		SINFRUP00000155363	scaffold_716	35,269-39,611.
Fugu		SINFRUP00000157228	scaffold_305	44,870-46,603.
Fugu		SINFRUP00000170718	scaffold_305	41,858-42,676.
Fugu		SINFRUP00000164123	scaffold_462	117,948-121,088.
Xenopus		ENSXETP00000001218	scaffold_130	528,586-529,670.
Xenopus		GENSCAN00000027135	scaffold_48	2,631,405-2,632,370
Xenopus	NP_989039.1	ENSXETP00000023462	scaffold_277	1,183,627-1,184,676.
Xenopus		GENSCAN00000028902	scaffold_556	23,850-31,057.
Xenopus		ENSXETP00000044584	scaffold_713	529,705-530,571.
Xenopus	NP_001005433.1	ENSXETP00000047746	scaffold_92	404,770-405,711.
Xenopus		ENSXETP00000053971	scaffold_277	1,191,953-1,192,360.
Xenopus	AAI21412.1	ENSXETP00000057692	scaffold_50	3,698,831-3,699,916.
Xenopus		ENSXETP00000006574	scaffold_55	1,602,120-1,603,226.
Xenopus		ENSXETP00000014700	scaffold_617	169,876-181,095.
Xenopus		ENSXETP00000023463	scaffold_277	1,281,118-1,282,056.
Xenopus	NP_001008003.1	ENSXETP00000028210	scaffold_478	164,090-165,151.
Xenopus	NP_001025633.1	ENSXETP00000050144	scaffold_117	1,200,375-1,202,006.
Xenopus		ENSXETP00000056279	scaffold_55	1,659,665-1,704,222.
Xenopus		ENSXETP00000056282	scaffold_55	1,612,669-1,644,858.
Xenopus		ENSXETP00000022341	scaffold_258	869,068-870,102.
Xenopus		ENSXETP00000024345	scaffold_390	661,523-663,920.
Xenopus		ENSXETP00000041128	scaffold_877	291,943-292,974.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Xenopus		ENSXETP00000044651	scaffold_713	576,320-577,297.
Xenopus		ENSXETP00000053350	scaffold_258	688,808-689,779.
Xenopus		GENSCAN00000027075	scaffold_48	2,648,100-2,649,110.
Xenopus		GENSCAN00000027116	scaffold_48	2,640,321-2,641,340.
Xenopus	AAI21581.1	ENSXETP00000036876	scaffold_409	480,582-481,628.
Xenopus		ENSXETP00000056732	scaffold_49	2,548,951-2,551,091.
Xenopus	NP_001004810.1	ENSXETP00000036888	scaffold_409	472,761-473,795.
Xenopus		GENSCAN00000056936	scaffold_258	721,618-722,670.
Xenopus		ENSXETP00000005770	scaffold_65	2,996,388-2,997,427.
Xenopus		ENSXETP00000029577	scaffold_94	297,797-301,454.
Xenopus		ENSXETP00000053195	scaffold_64	3,012,135-3,072,453.
Xenopus		ENSXETP00000007235	scaffold_113	2,609,822-2,623,507.
Xenopus		ENSXETP00000007247	scaffold_113	2,540,085-2,543,035.
Xenopus		ENSXETP00000037231	scaffold_679	513,772-514,803.
Xenopus		ENSXETP00000055469	scaffold_351	404,849-446,304.
Xenopus		ENSXETP00000055467	scaffold_351	561,182-586,791.
Xenopus		ENSXETP00000055471	scaffold_351	404,825-446,304.
Xenopus		ENSXETP00000055468	scaffold_351	445,507-561,179.
Xenopus		ENSXETP00000030483	scaffold_185	1,199,131-1,200,124.
Xenopus		ENSXETP00000035060	scaffold_284	533,931-534,971.
Xenopus		ENSXETP00000053184	scaffold_518	693,316-694,374
Xenopus	NP_001025605.1	GENSCAN00000014227	scaffold_15	2,143,102-2,157,196.
Xenopus		ENSXETP00000039815	scaffold_258	95,871-96,999.
Xenopus		ENSXETP00000054789	scaffold_35	815,812-816,315
Xenopus		ENSXETP00000044545	scaffold_713	241,894-266,458.
Xenopus		ENSXETP00000044559	scaffold_713	279,493-298,555.
Xenopus		ENSXETP00000044560	scaffold_713	310,839-311,852.
Xenopus		ENSXETP00000054000	scaffold_1836	17,997-20,696.
Xenopus		FGENESH00000006348	scaffold_689	106,939-107,796.
Xenopus		GENSCAN00000076358.	scaffold_702	381,863-384,168.
Xenopus		ENSXETP00000017831	scaffold_279.	976,995-982,811.
Xenopus		GENSCAN00000062538	scaffold_408	117,883-118,941.
Xenopus		GENSCAN00000053024	scaffold_31	3,188,819-3,211,089.
Xenopus		GENSCAN00000051641	scaffold_224	1,372,416-1,373,420.
Xenopus		ENSXETP00000044537	scaffold_713	241,363-242,793.
Xenopus		ENSXETP00000026874	scaffold_2462.	143-703.
Xenopus		ENSXETP00000039948	scaffold_258.	689,790-722,625.
Xenopus		ENSXETP00000037521	scaffold_556.	30,089-50,629.
Xenopus		ENSXETP00000023462	scaffold_277.	1,183,627-1,184,676.
Xenopus		ENSXETP00000037509	scaffold_91.	2,737,577-2,757,895.
Xenopus		GENSCAN00000040353.	scaffold_277.	1,427,223-1,463,817.

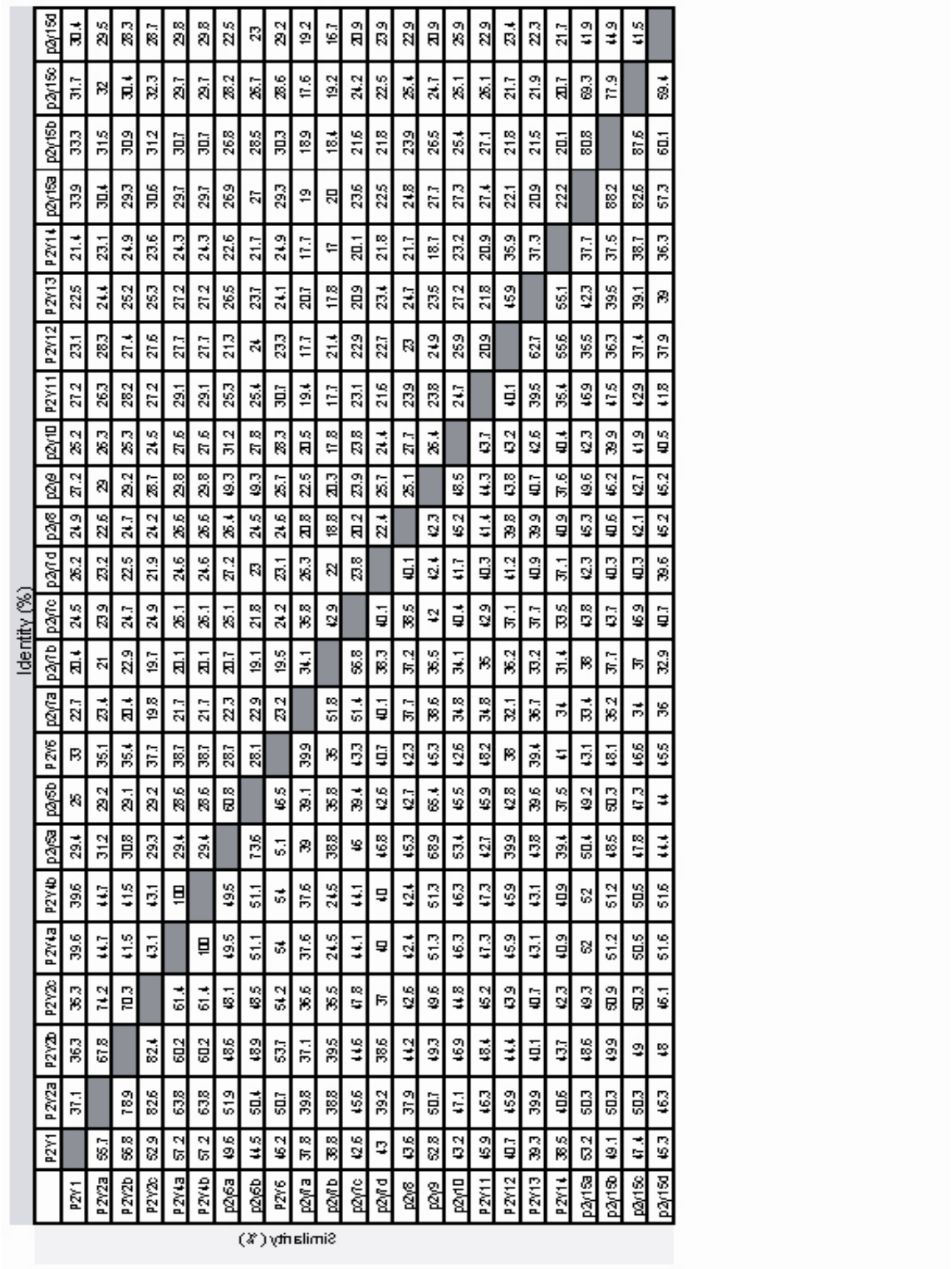
Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Human		ENSP00000171757	X	78,102,674-78,103,693.
Human		ENSP00000276077	X	78,313,161-78,314,162.
Human		ENSP00000309771	11	72,685,212-72,686,198.
Human		ENSP00000323872	19	10,083,398-10,086,414.
Human		ENSP00000344353	13	47,883,526-47,884,560.
Human		ENSP00000362398	X	77,897,023-77,898,135.
Human		ENSP00000365888	13	96,437,001-96,438,014.
Human		ENSP00000258380	2	231,482,962-231,483,921.
Human		ENSP00000301974	1	28,349,091-28,350,119.
Human		ENSP00000310305	11	72,622,853-72,623,986.
Human		ENSP00000327875	12	6,599,557-6,600,675.
Human		ENSP00000355156	3	153,080,371-153,082,026.
Human		ENSP00000362401	X	77,414,886-77,415,899.
Human		ENSP00000272644	2	128,124,057-128,125,799.
Human		ENSP00000282018	13	48,178,955-48,179,995.
Human		ENSP00000304767	3	154,036,262-154,037,383.
Human		ENSP00000322731	2	241,218,043-241,218,972.
Human		ENSP00000335289	X	69,395,102-69,396,199.
Human		ENSP00000362397	X	78,227,089-78,228,101.
Human		ENSP00000365596	13	98,745,315-98,746,400.
Human		ENSP00000372385	3	160,041,482-160,042,567.
Human		ENSP00000238699	14	90,770,050-90,771,177.
Human		ENSP00000328818	14	104,588,376-104,592,802.
Human		ENSP00000267549	14	87,546,945-87,547,958.
Human		ENSP00000319744	19	50,785,876-50,786,964.
Human		ENSP00000248076	19	16,861,001-16,862,432.
Human		ENSP00000321326	5	76,047,889-76,065,084.
Human		ENSP00000296641	5	75,949,163-75,954,811.
Human		ENSP00000296677	5	76,150,764-76,165,382.
Human		ENSP00000370697	X	1,544,372-1,545,451.
Human		ENSP00000307713	14	95,773,198-95,777,594.
Human		ENSP00000246538	19	40,554,102-40,555,142.
Human		ENSP00000246549	19	40,632,457-40,633,449.
Human		ENSP00000328230	19	40,541,633-40,542,673.
Human		ENSP00000216629	14	95,799,773-95,800,834
Human		ENSP00000273430	3	149,941,513-149,942,592.
Human		ENSP00000360973	X	115,217,562-115,218,653.
Human		ENSP00000367376	X	41,439,831-41,440,976.
Human		ENSP00000307259	3	152,538,295-152,539,323.
Human		ENSP00000308479	3	152,398,904-152,399,863.
Human		ENSP00000308361	3	152,413,778-152,414,794.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Human		ENSP00000260843	3	152,494,647-152,500,578.
Human		ENSP00000320376	3	152,528,469-152,529,470
Human		ENSP00000307445	14	23,854,698-23,855,756.
Lamprey		GENSCAN00000158257	Contig Contig28082.1	1,432-4,885.
Lamprey		GENSCAN00000095337	Contig1237.5	6,089-8,552.
Lamprey		GENSCAN00000077202	Contig1237.3	3,423-8,820.
Lamprey		GENSCAN00000136436	Contig1155.4	2,435-5,646.
Lamprey		GENSCAN00000084877	Contig23485.2	498-5,191.
Lamprey		GENSCAN00000157564	Contig20098.3	63-1,514.
Lamprey		GENSCAN00000121221	Contig39423.2	848-1,907.
Lamprey		GENSCAN00000156679	Contig12084.1	4,224-5,243.
Lamprey		GENSCAN00000037718	Contig12724.3	2,531-3,526.
Lamprey		GENSCAN00000044642	Contig4826.2	752-6,901.
Lamprey		GENSCAN00000063763	Contig40178.2	683-1,795.
Lamprey		GENSCAN00000009189	Contig51058.2	818-1,990.
Lamprey		GENSCAN00000146267	Contig20197.3	2,582-3,661.
Lamprey		GENSCAN00000119553	Contig17584.5	2,999-4,018.
Lamprey		GENSCAN00000128815	Contig23839.5	651-1,382.
Lamprey		GENSCAN00000020987	Contig4102.4	8,105-9,136.
Lamprey		GENSCAN00000047780	Contig16137.2	3,192-5,482.
Lamprey		GENSCAN00000036578	Contig6216.2	8,939-10,219.
Lamprey		GENSCAN00000094439	Contig5757.3	6,568-7,674.
Lamprey		GENSCAN00000007934	Contig97.6	1,329-2,297.
Lamprey		GENSCAN00000036479	Contig1237.4	44-528.
Lamprey		GENSCAN00000111041	Contig80603.2	994-1,854.
Lamprey		GENSCAN00000036489	Contig15747.3	2,752-3,891.
Lamprey		GENSCAN00000011025	Contig28634.4	2,445-3,218.
Lamprey		GENSCAN00000042303	Contig1593.6	6,057-11,008.
Lamprey		GENSCAN00000014373	Contig905.7	12,484-13,392.
Lamprey		GENSCAN00000053656	Contig6884.5	1,754-3,814.
Lamprey		GENSCAN00000005865	Contig1100.6	1,049-2,300.
Lamprey		GENSCAN00000119531	Contig7476.7	2,776-3,969.
Lamprey		GENSCAN00000035238	Contig26957.2	197-1,066.
Lamprey		GENSCAN00000118369	Contig9287.1	5,108-6,220.
Lamprey		GENSCAN00000157983	Contig571.8	2,975-6,823.
Lamprey		GENSCAN00000025544	Contig25590.1	314-1,366.
Lamprey		GENSCAN00000095106	Contig20976.4	1,912-2,926.
Lamprey		GENSCAN00000046309	Contig1126.1	507-1,202.
Lamprey		GENSCAN00000035678	Contig80717.2	20-689.
Lamprey		GENSCAN00000135946	Contig76467.2	524-1,061.
Lamprey		GENSCAN00000135681	Contig17687.3	1,238-1,706.

Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Lamprey		GENSCAN00000027224	Contig905.5	3,391-4,299.
Lamprey		GENSCAN00000067925	Contig32517.2	472-1,793.
Lamprey		GENSCAN00000127875	Contig47802.2	24-972.
Lamprey		GENSCAN00000041420	Contig8376.4	504-1,742.
Lamprey		GENSCAN00000074896	Contig83438.1	570-1,520.
Lamprey		GENSCAN00000136806	Contig7537.2	432-7,942.
Lamprey		GENSCAN00000091571	Contig11380.5	17-910.
Lamprey		GENSCAN00000043020	Contig7879.4	6,931-8,169.
Lamprey		GENSCAN00000065760	Contig1968.8	198-1,721.
Lamprey		GENSCAN00000151939	Contig1968.7	2,616-3,830.
Lamprey		GENSCAN00000015235		551-1,678.
Lamprey		GENSCAN00000097352	Contig1968.9	3,650-6,836.
Lamprey		GENSCAN00000089907	Contig3453.6	1,109-4,266.
Lamprey		GENSCAN00000036448	Contig1593.1	1,687-2,637.
Lamprey		GENSCAN00000054943	Contig17736.3	3,495-4,436.
Lamprey		GENSCAN00000014065	Contig2020.1	19,415-20,392.
Elephant shark	AAVX01000008			1099306860870
Elephant shark	AAVX01005288			1099306865997
Elephant shark	AAVX01011008			1099306871677
Elephant shark	AAVX01016502			1099306877170
Elephant shark	AAVX01025604			1099306886267
Elephant shark	AAVX01026588			1099306887249
Elephant shark	AAVX01028503			1099306889164
Elephant shark	AAVX01051690			1099306912159
Elephant shark	AAVX01056054			1099306916507
Elephant shark	AAVX01056829			1099306917278
Elephant shark	AAVX01059132			1099306919568
Elephant shark	AAVX01062489			1099306922907
Elephant shark	AAVX01064747			1099306925154
Elephant shark	AAVX01065781			1099306926180
Elephant shark	AAVX01072944			1099306933185
Elephant shark	AAVX01074580			1099306933605
Elephant shark	AAVX01096862			1099306955612
Elephant shark	AAVX01103357			1099306962107
Elephant shark	AAVX01109705			1099306968455
Elephant shark	AAVX01123176			1099306981927
Elephant shark	AAVX01124620			1099306983371
Elephant shark	AAVX01130182			1099306988934
Elephant shark	AAVX01132629			1099306991381
Elephant shark	AAVX01153499			1099306385629
Elephant shark	AAVX01198344			1099306506645

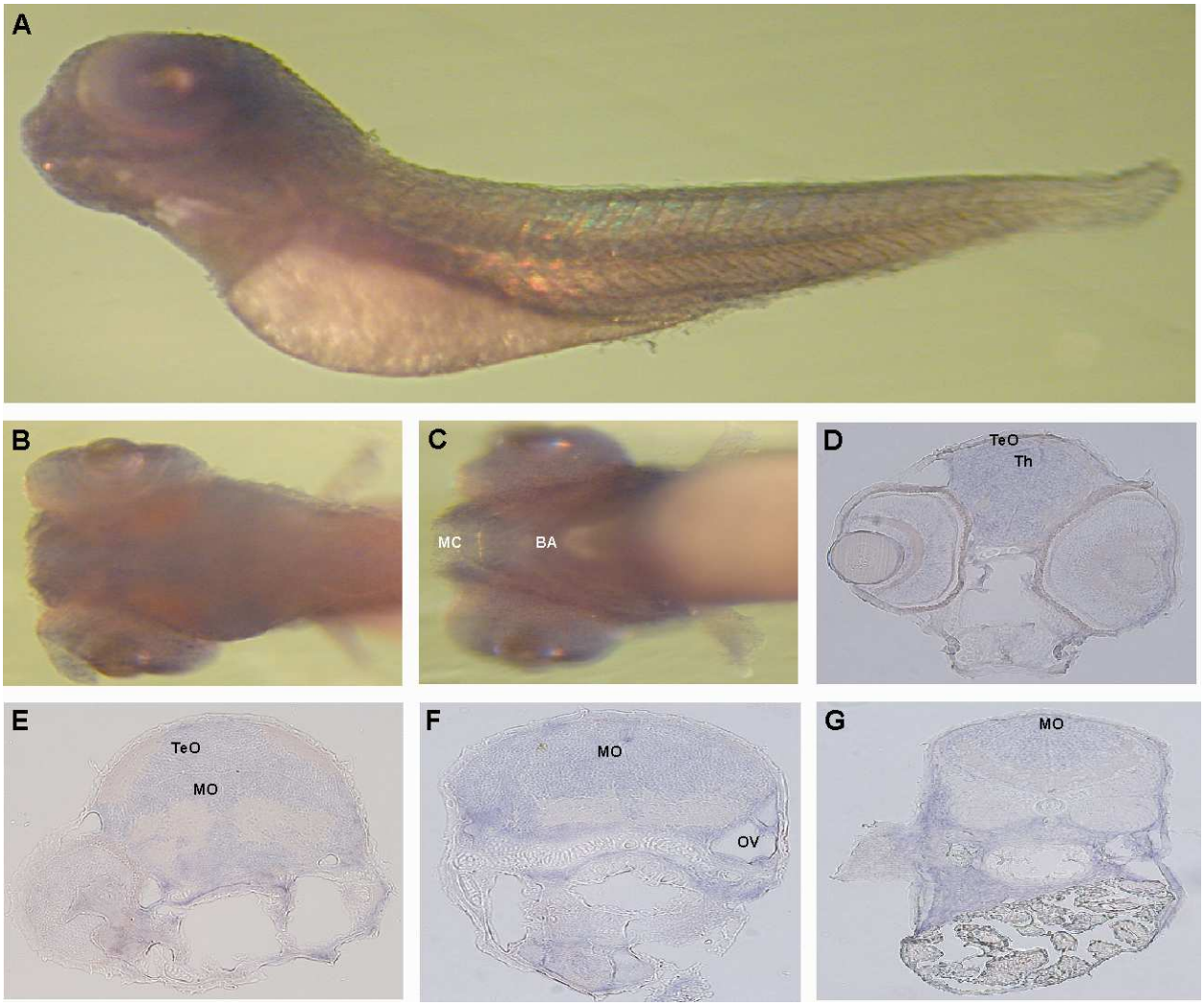
Species	NCBI	Ensembl protein database	Chromosome/Scaffold	Location
Elephant shark	AAVX01206512			1099306528802
Elephant shark	AAVX01207006			1099306530160
Elephant shark	AAVX01234747			1099306605966
Elephant shark	AAVX01247055			1099306639703
Elephant shark	AAVX01292903			1099306763742
Elephant shark	AAVX01326185			1099306851145
Elephant shark	AAVX01373377			1236908_1
Elephant shark	AAVX01395879			1098890040701
Elephant shark	AAVX01444803			1098980708334
Elephant shark	AAVX01500497			1098882303600
Little Skate	AAG42684			

C. Pairwise comparison of zebrafish P2Y genes



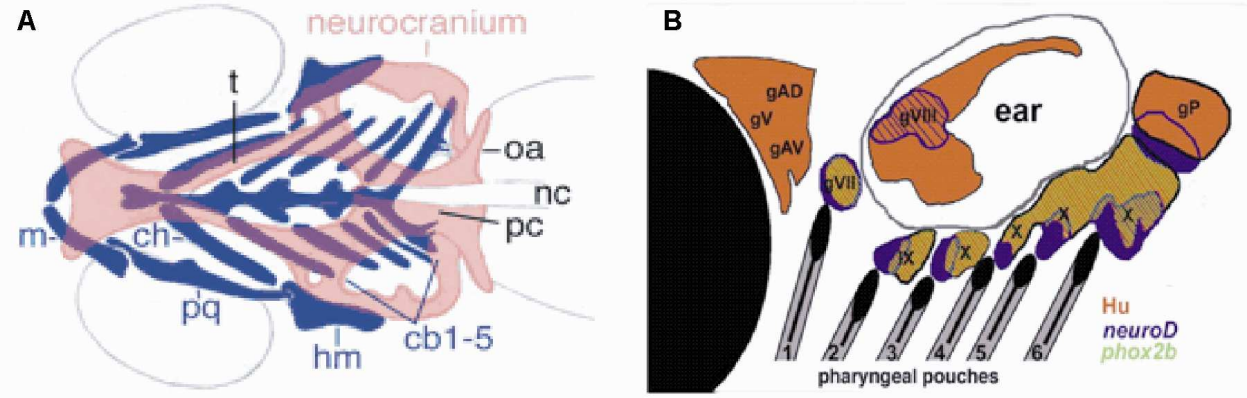
	P2Y1	P2Y2a	P2Y2b	P2Y2c	P2Y4a	P2Y4b	p2y3a	p2y5b	P2Y6	p2y7a	p2y7b	p2y7c	p2y7d	p2y8	p2y9	p2y10	P2Y11	P2Y12	P2Y13	P2Y14	p2y15a	p2y15b	p2y15c	p2y15d
P2Y1		37.1	36.3	35.3	39.6	39.6	29.4	25	33	22.7	20.4	24.5	26.2	24.9	27.2	25.2	27.2	23.1	22.5	21.4	33.9	33.3	31.7	30.4
P2Y2a	65.7		67.8	74.2	44.7	44.7	31.2	29.2	35.1	23.4	21	23.9	23.2	22.6	24.9	29	26.3	28.3	24.4	23.1	30.4	31.5	32	29.5
P2Y2b	66.8	78.9		70.3	41.5	41.5	30.8	29.1	35.4	20.4	22.9	24.7	22.5	24.7	29.2	25.3	28.2	27.4	25.2	24.9	29.3	30.9	30.4	28.3
P2Y2c	62.9	82.6	82.4		43.1	43.1	29.3	29.2	37.7	19.8	19.7	24.9	21.9	24.2	28.7	24.5	27.2	27.6	25.3	23.6	30.6	31.2	32.3	28.7
P2Y4a	67.2	63.8	60.2	61.4	100		29.4	28.6	38.7	21.7	20.1	26.1	24.6	26.6	29.8	27.6	29.1	27.7	27.2	24.3	29.7	30.7	29.7	29.8
P2Y4b	67.2	63.8	60.2	61.4	100	100		28.6	38.7	21.7	20.1	26.1	24.6	26.6	29.8	27.6	29.1	27.7	27.2	24.3	29.7	30.7	29.7	29.8
p2y6a	49.6	51.9	48.6	48.1	49.5	49.5		60.8	28.7	22.3	20.7	25.1	27.2	26.4	29.3	31.2	25.3	21.3	26.5	22.6	26.9	26.8	28.2	22.5
p2y6b	44.5	50.4	48.9	48.5	51.1	51.1	73.6		28.1	22.9	19.1	21.8	23	24.5	49.3	27.8	25.4	24	23.7	21.7	27	28.5	26.7	23
P2Y6	46.2	50.7	53.7	54.2	54	54	51.1	46.5		23.2	19.5	24.2	23.1	24.6	25.7	28.3	30.7	23.3	24.1	24.9	29.3	30.3	28.6	29.2
p2y7a	37.8	39.8	37.1	36.6	37.6	37.6	39	39.1	39.9		34.1	35.8	26.3	20.8	22.5	20.5	19.4	17.7	20.7	17.7	19	18.9	17.6	19.2
p2y7b	38.8	38.8	39.5	36.5	24.5	24.5	38.8	36.8	36	51.8		42.9	22	18.8	20.3	17.8	17.7	21.4	17.8	17	20	18.4	19.2	16.7
p2y7c	42.6	45.6	44.6	47.8	44.1	44.1	46	39.4	43.3	51.4	56.8		23.8	20.2	23.9	23.8	23.1	22.9	20.9	20.1	23.6	21.6	24.2	20.9
p2y7d	43	39.2	38.6	37	40	40	46.8	42.6	40.7	40.1	38.3	40.1		22.4	25.1	24.4	21.6	22.7	23.4	21.8	22.5	21.8	22.5	23.9
p2y8	43.6	37.9	44.2	42.6	42.4	42.4	45.3	42.7	42.3	37.7	37.2	38.5	40.1		25.1	27.7	23.9	23	24.7	21.7	24.8	23.9	25.4	22.9
p2y9	52.8	50.7	49.3	49.6	51.3	51.3	68.9	65.4	45.3	38.6	36.5	42	42.4	42.3		26.4	23.8	24.9	23.5	18.7	27.7	26.5	24.7	20.9
p2y10	43.2	47.1	46.9	44.8	46.3	46.3	53.4	45.5	42.6	34.8	34.1	40.4	41.7	45.2	48.5		24.7	25.9	27.2	23.2	27.3	25.4	25.1	25.9
P2Y11	45.9	46.3	48.4	45.2	47.3	47.3	42.7	45.9	48.2	34.8	35	42.9	40.3	41.4	44.3	43.7		20.9	21.8	20.9	27.4	27.1	26.1	22.9
P2Y12	40.7	45.9	44.4	43.9	45.9	45.9	39.9	42.8	38	32.1	36.2	37.1	41.2	39.8	43.8	43.2	40.1		45.9	35.9	22.1	21.8	21.7	23.4
P2Y13	39.3	39.9	40.1	40.7	43.1	43.1	43.8	39.6	39.4	36.7	33.2	37.7	40.9	39.9	40.7	42.6	39.5	62.7		37.3	20.9	21.5	21.9	22.3
P2Y14	38.5	40.6	43.7	42.3	40.9	40.9	39.4	37.5	41	34	31.4	33.5	37.1	40.9	37.6	40.4	35.4	55.6	55.1		22.2	20.1	20.7	21.7
p2y15a	63.2	60.3	49.3	62	62	62	60.4	49.2	43.1	33.4	38	43.8	42.3	45.3	49.6	42.3	45.9	35.5	42.3	37.7		80.8	69.3	61.9
p2y15b	49.1	50.3	50.9	51.2	51.2	51.2	48.5	50.3	48.1	36.2	37.7	43.7	40.3	40.6	46.2	39.9	47.5	36.3	39.5	37.5	88.2		77.9	64.9
p2y15c	47.4	50.3	49	50.3	50.5	50.5	47.8	47.3	46.6	34	37	46.9	40.3	42.1	42.7	41.9	42.9	37.4	39.1	38.7	82.6	87.6		61.5
p2y15d	45.3	46.3	48	46.1	51.6	51.6	44.4	44	45.5	36	32.9	40.7	39.6	45.2	45.2	40.5	41.8	37.9	39	36.3	57.3	60.1	59.4	

D. Expression of FFAR2c in brain and pharyngeal arches.



Analysis of FFAR2c expression in 5dpf larval by whole mount in situ hybridization. Panel (A-C), whole mount and (D-G) transverse sections after hybridization. A) Lateral view shows expression in the brain and pharyngeal arches; (B) Dorsal view shows expression in the brain; (C) Ventral view shows expression in Merkel's cartilage and branchial arches; In the brain, the expression is distributed from telencephalon to medulla oblongata in rhombencephalon (D-F); Expression is also detected in otic vesicle (F). *Abbreviations*: BA: branchial arches; MC: Merkel's cartilage; MO: medulla oblongata; OV: otic vesicle; TeO: optic tectum; Th: thalamus; VG: vagal ganglion.

E. Schematic drawing of pharyngeal arches of zebrafish



A) Schematic drawing of cartilage in ventral view of 5 dpf larval adapted from Piotrowski *et al*., 2003. B) Lateral view of pharyngeal branches in zebrafish adapted from Holzschuh *et al*., 2005. *Abbreviations*: Cb 1-5, ceratobranchial cartilages; ch, ceratohyal cartilage; gAD, dorsal anterior lateral line; gAV, ventral anterior lateral line; gP, posterior lateral line; gV, trigeminal; gVII, facial (geniculate), gVIII, auditory; gIX, glossopharyngeal (petrosal); gX, vagal (nodose); hm, hyomandibula; m, Meckel's cartilage of mandibular arch; n, notochord; oa, occipital arch; pq, palatoquadrate; pc, parachordalia; t, trabeculae.

8. CURRICULUM VITAE

Name:	Yen Yen, Kwan
Date of birth:	July 12, 1978
Nationality:	Malaysian
Address:	Apt 66, Zuelpicherstr. 58e 50674 Cologne, Germany
Phone:	+49 (0176) 2050 2746

EDUCATION

Sept 2004 – Present	University of Cologne, Germany PhD scholarship holder of the International Graduate School In Genetics and Functional Genomics Project: Phylogenomic analysis of metabotropic P2Y receptor family and Its expression in zebrafish, *Danio rerio* Supervisor: Prof Dr Sigrun Korsching
Nov 2001 – Sept 2004	University of Putra, Malaysia MSc in Genetic Engineering and Molecular Biology Recipient of Malaysian Oil Palm Board Graduate Programme Project: A catalogue of oil palm ESTs from a floral cDNA library and the analysis of two MADS box genes Supervisor: Dr Sharifah Shahrul Rabiah Syed Alwee
May 1998 – July 2001	University of Putra, Malaysia BSc in Biotechnology Project: The construction and screening of low abundance genes from vegetative meristem cDNA library of oil palm Supervisor: Dr Harikrishna Kulaveerasingam

WORK EXPERIENCE

Jan 2004 – Aug 2004	Department of Biotechnology, University of Putra, Malaysia Group leader: Dr Ho Chai Ling Research assistant Project: Gene expression of carbohydrate metabolic pathway in oil palm in response to environmental stresses

PUBLICATIONS

The fifth class of G alpha proteins
Yuichiro Oka, Luis Saraiva, **Yen Yen Kwan** and Sigrun Korsching
PNAS 2009, 106(5):1484-9

Analysis and functional annotation of expressed sequence tags (ESTs) from multiple tissues of oil palm (*Elaeis quineedsis* Jacq.)
Ho CL, **Kwan YY**, Choi MC, Tee SS, Ng WH, Lim KA, Lee YP, Ooi SE, Lee WW, Tee JM, Tan SH, Kulaveerasingam H, Alwee SSRS and Abdullah MO
BMC Genomics 2007, 8:381